

Industrial

Electronics

Biomedical

Civil

Aerospace

Computer

Electrical

Chemical

Mechanical



ENGINEERING



TIME TO CHANGE THE WAY ACADEMICS WORKS? TRY ACAMEDICS!

If you are not using the Internet, please skip this ad and enjoy the rest of the journal. The average burden of reading this ad is two minutes, which may add two months of productivity to your academic life. Seriously!

Whether you organize a conference, publish a journal, or serve on a committee to collect and review applications, you can use the Internet to make your life easier. We are not talking about emails. We are talking about state-of-the-art online systems to collect submissions, assign them to reviewers, and finally make a decision about each submission. We are talking about value-added services, such as payment and registration systems to collect registration fees online. We are talking about digital document services, such as proceedings CD/DVD development and duplication, or creating professional-looking digital documents.

Finally, we are talking about AFFORDABLE PRICE, QUALITY, and CUSTOMIZATION. And we are talking about each of them at the same time.

By the way, you don't have to be a computer geek to use our systems. We have a couple of them, and they will do all the technical mumbo jumbo for you. We also have a few select people from academics like you, and they know what you do. You just relax and enjoy our systems.

If you are still reading this ad, chances are you are interested in our systems or services. So, visit us at www.acamedics.com. While you are there, check the names of our clients as well. Most of them are quite familiar, but the list is too long to include here.



INTERNATIONAL JOURNAL OF MODERN ENGINEERING

The INTERNATIONAL JOURNAL OF MODERN ENGINEERING (IJME) is an independent, not-for-profit publication, which aims to provide the engineering community with a resource and forum for scholarly expression and reflection.

IJME is published twice annually (Fall and Spring issues) and includes peer-reviewed articles, book and software reviews, editorials, and commentary that contribute to our understanding of the issues, problems, and research associated with engineering and related fields. The journal encourages the submission of manuscripts from private, public, and academic sectors. The views expressed are those of the authors and do not necessarily reflect the opinions of IJME or its editors.

EDITORIAL OFFICE:

Mark Rajai, Ph.D.
Editor-in-Chief
Office: (818) 677-2167
Email: ijmeeditor@iajc.org
Dept. of Manufacturing Systems
Engineering & Management
California State University
Northridge
18111 Nordhoff Street
Northridge, CA 91330-8332

THE INTERNATIONAL JOURNAL OF MODERN ENGINEERING EDITORS

Editor-in-Chief:

Mark Rajai

California State University-Northridge

Associate Editors:

Alok Verma

Old Dominion University

Li Tan

Purdue University North Central

Production Editor:

Philip Weinsier

Bowling Green State University-Firelands

Subscription Editor:

Morteza Sadat-Hossieny

Northern Kentucky University

Financial Editor:

Li Tan

Purdue University North Central

Executive Editor:

Sohail Anwar

Penn State University

Manuscript Editor:

Philip Weinsier

Bowling Green State University-Firelands

Copy Editors:

Victor J. Gallardo

University of Houston

Li Tan

Purdue University North Central

Publishers:

Jerry Waite

University of Houston

Hisham Alnajjar

University of Hartford

Web Administrator:

Saeed Namyar

Namyar Computer Solutions

TABLE OF CONTENTS

<i>Editor's Note: Upcoming IAJC-ASEE Joint International Conference</i>	3
<i>Philip Weinsier, IJME Manuscript Editor</i>	
<i>Second Order task Specifications in the Geometric design of Spatial Mechanical Linkages</i>	5
<i>Nina P. Robson, Texas A&M University; J. Michael McCarthy, University of California, Irvine</i>	
<i>Application of Six Sigma to Gear Box Manufacturing</i>	12
<i>Kambiz Farahmand, North Dakota State University; Jose Victorino Marquez Grajales; Mohsen Hamidi, North Dakota State University</i>	
<i>Microstepping Control of Stepper Motors with Data Interpolation and Direct Voltage Control</i>	20
<i>Dong-Hee Lee, Kyungshung University; Shiyoung Lee, the Penn State University Berks Campus; Tae-Hyun Won, Dongeui Institute of Techology</i>	
<i>An Enhanced Framework for Multi-module Embedded Reconfigurable Systems</i>	27
<i>Muhammad Z. Hasan, Texas A & M University; Timothy Davis, Texas A & M University; Troy Kensinger, Texas A & M University; Sotirios G. Ziavras, New Jersey Institute of Technology</i>	
<i>Sound Source Localization Employing Polar Directivity Patterns of Bidirectional Microphones</i>	35
<i>Vijay Varada, The University of Toledo; Hong Wang, The University of Toledo; Vijay Devabhaktuni, The University of Toledo</i>	
<i>Modeling and Validation of Autonomous Rendezvous and Docking of Air Bearing Vehicles</i>	45
<i>Amir Mobasher, Alabama A&M University; Paul Shiue; Christian Brothers University; Hossein Jamshidi, Alabama A&M University; Zhengtao Deng, Alabama A&M University</i>	
<i>Case Study of an Adaptive Automated Health Insurance Fraud Auditor</i>	52
<i>Fletcher Lu, University of Ontario Institute of Technology</i>	
<i>A Neural Oscillator Model for Tinnitus and its Management by Sound Therapy</i>	58
<i>Hirofumi Nagashino, The University of Tokushima; Ken'ichi Fujimoto, The University of Tokushima; Yohsuke Kinouchi, The University of Tokushima; Ali A. Danesh, Florida Atlantic University; Abhijit S. Pandya, Florida Atlantic University</i>	
<i>Hydroelectric Plant Design as a Synthesis Tool for an Engineering Technology Curriculum</i>	67
<i>Henry Foust, Nicholls State University; George Watt, Nicholls State University</i>	
<i>Computation of Shockwave Structures in Weakly Ionized Gases by Solving Burnett and Modified Rankine-Hugoniot Equations</i>	74
<i>Xiaoqing Qian, Z.T. Deng, Alabama A&M University</i>	
<i>Interactive Data Visualization and Analysis for Mobile-Phone Performance Evaluation</i>	82
<i>Yongsuk Lee, Florida Atlantic University (FAU), Xingquan Zhu, FAU, Abhijit Pandya, FAU, and Sam Hsu, FAU</i>	
<i>Instructions for Authors: Manuscript Requirements</i>	95

EDITOR'S NOTE: UPCOMING IAJC-ASEE JOINT INTERNATIONAL CONFERENCE



Philip Weinsier, IJME Manuscript Editor

IAJC-ASEE 2011 Joint International Conference

The editors and staff at IAJC would like to thank you, our readers, for your continued support and look forward to seeing you at the upcoming IAJC conference. For this third biennial IAJC conference, we will be partnering with the American Society for Engineering Education (ASEE). This event will be held at the University of Hartford, CT, April 29-30, 2011, and is sponsored by IAJC, ASEE and IEEE (the Institute of Electrical and Electronic Engineers).

The IAJC-ASEE Conference Committee is pleased to invite faculty, students, researchers, engineers, and practitioners to present their latest accomplishments and innovations in all areas of engineering, engineering technology, math, science and related technologies.

Presentation papers selected from the conference will be considered for publication in one of the three IAJC journals or other affiliate journals. Oftentimes, these papers, along with manuscripts submitted at-large, are reviewed and published in less than half the time of other journals. Please refer to the publishing details at the back of this journal, or visit us at www.iajc.org, where you can also read any of our previously published journal issues, as well as obtain information on chapters, membership and benefits, and journals.

IAJC Welcomes Three New Affiliate Journals

IAJC, the parent organization of the *International Journal of Modern Engineering* (IJME), the *International Journal of Engineering Research and Innovation* (IJERI) and the *Technology Interface International Journal* (TIJ), is a first-of-its-kind, pioneering organization acting as a global, multi-layered umbrella consortium of academic journals, conferences, organizations, and individuals committed to advanc-

ing excellence in all aspects of education related to engineering and technology. IAJC is fast becoming the association of choice for many researchers and faculty due to its high standards, personal attention, fast-track publishing, biennial IAJC conferences, and its diversity of journals.

In 2010, IAJC accepted the Technology Interface International Journal as the third official, IAJC-owned journal. Also welcomed to the growing list of affiliate journals are the *International Journal of Engineering* (IJE), the *International Journal of Industrial Engineering Computations* (IJIEC) and the *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies* (ITJEMAST). With three official IAJC-owned journals and 10 affiliate journals, authors now have a venue for publishing work across a broad range of topics.

Current Issue of IJME

The acceptance rates for IJME range from about 20-45%. This issue saw an abundance of quality papers; thus, the acceptance rate for this issue was roughly 45%. And, due to the hard work of the IJME editorial review board, I am confident that you will appreciate the articles published here. IJME, along with IJERI and TIJ, are available online (www.ijme.us, www.ijeri.org & www.tij.org) and in print.

International Review Board

IJME is steered by IAJC's distinguished Board of Directors and is supported by an international review board consisting of prominent individuals representing many well-known universities, colleges, and corporations in the United States and abroad. To maintain this high-quality journal, manuscripts that appear in the *Articles* section have been subjected to a rigorous review process. This includes blind reviews by three or more members of the international editorial review board—with expertise in a directly related field—followed by a detailed review by the journal editors.

Acknowledgment

Listed here are the members of the editorial board, who devoted countless hours to the review of the many manuscripts that were submitted for publication. Manuscript reviews require insight into the content, technical expertise related to the subject matter, and a professional background in statistical tools and measures. Furthermore, revised manuscripts typically are returned to the same reviewers for a second review, as they already have an intimate knowledge of the work. So I would like to take this opportunity to thank all of the members of the review board.

Editorial Review Board Members

If you are interested in becoming a member of the IJME editorial review board, go to the IJME web site (Submissions page) and send me—Philip Weinsier, Manuscript Editor—an email.

Mehmet Bahadir	Murray State University	Zaki Kuruppallil	Ohio University (OH)
Rendong Bai	Eastern Illinois University	Ronald Land	Penn State University (PA)
Boris Blyukher	Indiana State University (IN)	Jay Lee	Purdue University Calumet (IN)
Walter Buchanan	Texas A&M University (TX)	Shiyong Lee	Penn State University Berks (PA)
Isaac Chang	Cal Poly State University SLO (CA)	Soo-Yen Lee	Central Michigan University (MI)
Hans Chapman	Morehead State University (KY)	Chao Li	Florida A&M University (FL)
Rigoberto Chinchilla	Eastern Illinois University (IL)	Jimmy Linn	Eastern Carolina University (NC)
Raj Chowdhury	Kent State University (OH)	G.H. Massiha	University of Louisiana (LA)
Michael Coffman	Southern Illinois University (IL)	Jim Mayrose	Buffalo State College (NY)
Kanchan Das	East Carolina University (NC)	Thomas McDonald	Eastern Illinois University (IL)
Brad Deken	Southeast Missouri State University (MO)	David Melton	Eastern Illinois University (IL)
Dave Dillon	North Carolina A&T State University	Sam Mryyan	Excelsior College (NY)
Mehran Elahi	Elizabeth City State University (NC)	Wilson Naik	University of Hyderabad (INDIA)
Ahmed Elsayy	Tennessee Tech University (TN)	Arun Nambiar	California State U.—Fresno (CA)
Bob English	Indiana State University (IN)	Argie Nichols	University Arkansas Fort Smith (AR)
Rasoul Esfahani	DeVry University, USA	Hamed Niroumand	University Teknologi Malaysia
Fereshteh Fatehi	North Carolina A&T State U. (NC)	Troy Ollison	University of Central Missouri (MO)
Verna Fitzsimmons	Kent State University (OH)	Reynaldo M Pablo Jr.	Indiana University - Purdue (IN)
Richard Freeman	US Coast Guard Academy (CT)	Basile Panoutsopoulos	United States Navy
Vladimir Genis	Drexel University (PA)	Jose Pena	Purdue University Calumet (MI)
Liping Guo	Northern Illinois University (IL)	Karl Perusich	Purdue University (IN)
Rita Hawkins	Missouri State University (MO)	Thongchai Phairoh	Virginia State University (VA)
Youcef Himri	Safety Engineer in SONELGAZ Algeria	John Rajadas	Arizona State University (AZ)
Xiaobing Hou	Central Connecticut State University (CT)	Desire Rasolomampionona	Warsaw U. of Technology (POLAND)
Shelton Houston	University of Louisiana at Lafayette (LA)	Mulchand Rathod	Wayne State University (MI)
Luke Huang	University of North Dakota (ND)	Sangram Redkar	Arizona State University-Poly (AZ)
Dave Hunter	Western Illinois University (IL)	Michael Reynolds	University Arkansas Fort Smith (AR)
Pete Hylton	Indiana University Purdue (IN)	Mehdi Safari	Isfahan University of Technology (IRAN)
Anwar Jawad	University of Technology Baghdad IRAQ	Anca Sala	Baker College (MI)
Rex Kanu	Ball State University (IN)	Hiral Shah	St. Cloud State University
Khurram Kazi	Acadiaoptronics (MD)	Ehsan Sheybani	Virginia State University (VA)
Ognjen Kuljaca	Alcorn State University (MS)	Carl Spezia	Southern Illinois University (IL)
Chakresh Kumar	Uttar Pradesh Technology Uni. (INDIA)	Randy Stein	Ferris State University (MI)
		Adam Stienecker	Ohio Northern University (OH)
		Pingping Sun	IBM
		Jalal Taheri	Bostan Abad Islamic Azad U. (IRAN)
		Li Tan	Purdue University North Central (IN)
		Li-Shiang Tsay	North Carolina Ag & Tech State (NC)
		Philip Waldrop	Georgia Southern University (GA)
		Liangmo Wang	Nanjing University of Science & Technology (CHINA)
			Sam Houston State University (TX)
		Faruk Yildiz	U. of Maryland Eastern Shore (MD)
		Emin Yilmaz	Jackson State University (MS)
		Pao-Chiang Yuan	Missouri Western State U. (MO)
		Jinwen Zhu	

Second Order Task Specifications in the Geometric Design of Spatial Mechanical Linkages

Nina P. Robson, Texas A&M University; J. Michael McCarthy, University of California, Irvine

Abstract

This paper builds on the authors' planar kinematic synthesis for contact task specifications and formulates the kinematic specification of the synthesis problem for spatial open-serial chains in which a desired acceleration of the end-effector is specified.

Applications of this research focus on the design of spatial linkages to maintain specified local motion. A recently developed failure recovery strategy of a general six-degree-of-freedom TRS robotic arm is discussed and some experimental set up and tests of the proposed failure recovery are presented. The authors also briefly show another possible application of the geometric design of linkages using acceleration task specifications. It combines the second-order effects of the task with the particular kinematics of the chain to yield free parameters that allow for more than one system to accomplish one and the same task.

Introduction

This study considered the synthesis of spatial chains to guide an end-effector through a number of multiply separated positions [1], [2]. The kinematic specification is a number of task positions with specified end-effector velocities and accelerations. The goal of this study was to obtain all of the solutions to a given task specification in order to design mechanical linkages that could move the end-effector smoothly through the specified task. Research in the synthesis of serial chains to achieve acceleration requirements is limited. It is primarily found in the synthesis theory for planar RR chains, and the work by Chen and Roth [3] for spatial chains. The use of second-order effects first appeared in the analysis of grasping in a work by Hanafusa and Asada [4], where planar objects are grasped with three elastic rods.

Cai and Roth [5] and Montana [6] developed an expression for the velocity of the point of contact between two rigid bodies that includes the curvature of the contact bodies. Second-order contact kinematics for regular contacts such as surface-surface, curve-curve, curve-surface and vertex-surface are formulated in a unified framework in the recent work of Park et al. [7], extending Montana's first-order contact kinematics for surface-surface contact only. Sarkar et al. [8] develop an expression for the acceleration of the contact point between two contacting bodies.

Second-order considerations have also appeared in work by Trinkle [9] in the study of stability of frictionless polyhedral objects in the presence of gravity. The mobility of bodies in contact has been studied using first-order theories that are based on notions of instantaneous force and velocities [10]. For example, Ohwovoriole and Roth [10] describe the relative motion of contacting bodies in terms of Screw Theory, which is a first-order theory. Using first-order notions, Reuleaux [11], Mishra et al. [12] and Markenskoff et al. [13], derive bounds on the number of frictionless point contacts required for force closure, which is one means of immobilizing an object. However, first-order theories are inadequate in practice. The source of deficiency is that the relative mobility of an object in contact with finger bodies is not an infinitesimal notion but a local one. One must consider the local motions of the object, not the tangential aspects of the motions, as employed by the first-order theories.

Rimon and Burdick [14], [15] show that acceleration properties of movement can be used to effectively constrain a rigid body for part-fixturing and grasping applications. In previous work by the authors, planar synthesis [16] was presented as a technique for deriving geometric constraints on position, velocity and acceleration from contact and curvature task requirements. These constraints yielded design equations that can be solved to determine the dimensions of the serial chain.

In this current study, the authors briefly present this planar approach, expand on the spatial approach [17], [18] and present some of the applications of second-order task specifications for the geometric design of spatial linkages.

Geometric Design of Planar Mechanical Linkages with Task Acceleration Specifications

Assume that the planar task consists of positioning an end-effector of a robot arm at a start and a finish position M^j , $j=1, \dots, n$, such that in these positions there are prescribed velocities and accelerations. Let the movement of a rigid body be defined by the parameterized set of 3×3 homogeneous transforms $[T(t)]=[R(t), \mathbf{d}(t)]$ constructed from a rotation matrix, $R(t)$, and translation vector $\mathbf{d}(t)$. A point \mathbf{p} fixed

in the moving body traces a trajectory $\mathbf{P}(t)$ in a fixed coordinate frame F such that:

$$\begin{Bmatrix} P_x(t) \\ P_y(t) \\ 1 \end{Bmatrix} = \begin{bmatrix} \cos\phi(t) & -\sin\phi(t) & d_x(t) \\ \sin\phi(t) & \cos\phi(t) & d_y(t) \\ 0 & 0 & 1 \end{bmatrix} \begin{Bmatrix} p_x \\ p_y \\ 1 \end{Bmatrix}, \quad (1)$$

or

$$\mathbf{P}(t) = [T(t)]\mathbf{p}. \quad (2)$$

The goal is to determine the movement of the end-effector as defined by $[T(t)]$.

The movement of M relative to a world frame F in the vicinity of a reference position, defined by $t=0$, can be expressed by the Taylor series expansion,

$$[T^j(t)] = [T_0^j] + [T_1^j]t + \frac{1}{2}[T_2^j]t^2 + \dots, \quad j = 1, \dots, n$$

where $[T_i^j] = \left. \frac{d^i[T^j]}{dt^i} \right|_{t=0}$. (3)

The matrices $[T_0^j]$, $[T_1^j]$ and $[T_2^j]$ are defined by the position, velocity and acceleration of the end-effector in the vicinity of each task position M^j . Therefore, a point \mathbf{p} in M has the trajectory $\mathbf{P}(t)$ defined by the equation

$$\mathbf{P}^j(t) = [T^j(t)]\mathbf{p} = [T_0^j + T_1^j t + \frac{1}{2}T_2^j t^2 + \dots]\mathbf{p}. \quad (4)$$

Let $\mathbf{p} = [T_0^j]^{-1}\mathbf{P}^j$, which yields

$$\begin{aligned} \mathbf{P}^j(t) &= [T_0^j + T_1^j t + \frac{1}{2}T_2^j t^2 + \dots][T_0^j]^{-1}\mathbf{P}^j, \\ &= [I + \Omega^j t + \frac{1}{2}\Lambda^j t^2 + \dots]\mathbf{P}^j, \end{aligned} \quad (5)$$

where

$$[\Omega^j] = \begin{bmatrix} 0 & -\phi_1 & d_{x1} + d_{x0}\phi_1 \\ \phi_1 & 0 & d_{y1} - d_{x0}\phi_1 \\ 0 & 0 & 0 \end{bmatrix}, \quad [\Lambda^j] = \begin{bmatrix} -\phi_1^2 & -\phi_2 & d_{x2} + d_{x0}\phi_1^2 + d_{x0}\phi_2 \\ \phi_2 & -\phi_1^2 & d_{y2} + d_{x0}\phi_1^2 - d_{x0}\phi_2 \\ 0 & 0 & 0 \end{bmatrix} \quad (6)$$

are the planar velocity and planar acceleration matrices, which are defined by the end-effector velocity and acceleration specifications in the vicinity of some task positions M^j , $j=1, \dots, n$.

For example, the design parameters for a planar RR chain are the coordinates $\mathbf{B}=(B_x, B_y)$ of the fixed pivot, the coordinates $\mathbf{P}^1=(P_x, P_y)$ of the moving pivot when the floating link is in the first position, and the length R of the link. In each task position the moving pivot \mathbf{P}^j is constrained to lie at the distance R from \mathbf{B} , so we have,

$$(\mathbf{P}^j(t) - \mathbf{B}) \cdot (\mathbf{P}^j(t) - \mathbf{B}) = R^2. \quad (7)$$

The first and second derivative of this equation provide the velocity constraint equation

$$\frac{d}{dt}\mathbf{P} \cdot (\mathbf{P} - \mathbf{B}) = 0, \quad (8)$$

and the acceleration constraint equation

$$\frac{d^2}{dt^2}\mathbf{P} \cdot (\mathbf{P} - \mathbf{B}) + \left(\frac{d}{dt}\mathbf{P}\right) \cdot \left(\frac{d}{dt}\mathbf{P}\right) = 0. \quad (9)$$

In order to determine the five design parameters, five design equations are required. Choosing one of the task positions to be the first and using the relative displacement matrices $[D_{1j}] = [T_0^j][T_0^1]^{-1}$ allow one to define coordinates \mathbf{P}^j taken by the moving pivot as follows:

$$\mathbf{P}^j = [D_{1j}]\mathbf{P}^1. \quad (10)$$

It is now possible to substitute \mathbf{P}^j in equation (7) to obtain

$$([D_{1j}]\mathbf{P}^1 - \mathbf{B}) \cdot ([D_{1j}]\mathbf{P}^1 - \mathbf{B}) = R^2, \quad i = 1, \dots, n. \quad (11)$$

These are the position design equations. Notice that $[D_{11}]$ is the 3 x 3 identity matrix. From our definition of the 3 x 3 velocity matrix, we have $\frac{d}{dt}\mathbf{P}^j = [\Omega_j][D_{1j}]\mathbf{P}^1$ and substituting \mathbf{P}^j into (8), we obtain the velocity design equations

$$([\Omega^j][D_{1j}]\mathbf{P}^1) \cdot ([D_{1j}]\mathbf{P}^1 - \mathbf{B}) = 0, \quad j = 1, \dots, n. \quad (12)$$

From our definition of the 3 x 3 acceleration matrix, we have $\frac{d^2}{dt^2}(\mathbf{P}^j) = [\Lambda_j][D_{1j}]\mathbf{P}^1$ and substituting \mathbf{P}^j in equation (9) yields

$$([\Lambda_j][D_{1j}]\mathbf{P}^1) \cdot ([D_{1j}]\mathbf{P}^1 - \mathbf{B}) + ([\Omega^j][D_{1j}]\mathbf{P}^1) \cdot ([\Omega^j][D_{1j}]\mathbf{P}^1) = 0, \quad (13)$$

where $j=1, \dots, n$. These are the acceleration design equations. Thus, for each of the n task positions, the position, velocity and acceleration design equations have the following form:

$$\begin{aligned} \mathcal{P}_j &: ([D_{1j}]\mathbf{P}^1 - \mathbf{B}) \cdot ([D_{1j}]\mathbf{P}^1 - \mathbf{B}) = R^2, \\ \mathcal{V}_j &: ([\Omega^j][D_{1j}]\mathbf{P}^1) \cdot ([D_{1j}]\mathbf{P}^1 - \mathbf{B}) = 0, \\ \mathcal{A}_j &: ([\Lambda^j][D_{1j}]\mathbf{P}^1) \cdot ([D_{1j}]\mathbf{P}^1 - \mathbf{B}) \\ &+ ([\Omega_j][D_{1j}]\mathbf{P}^1) \cdot ([\Omega_j][D_{1j}]\mathbf{P}^1) = 0, \quad j = 1, \dots, n. \end{aligned} \quad (14)$$

The algebraic solution to the set of four bilinear equations for an RR chain is presented in McCarthy [19] for the case of five position synthesis and applies without any changes to the design equations (14).

Geometric Design of Spatial Mechanical Linkages with Task Acceleration Specifications

Assume that the spatial task consists of positioning an end-effector of a robot arm at a start and a finish position M^j , $j=1, \dots, n$, such that in these positions there are prescribed velocities and accelerations. The rotation angles used to define the orientation of the moving body in space are chosen to be $(\theta_j, \Phi_j, \Psi_j)$, representing the longitude, latitude, and roll angles that position the z-axis of the moving frame in the j-th position. Thus, the rotation matrix $[A^j]$ is given by

$$[A^j] = [Y(\theta_j)][X(-\phi_j)][Z(\psi_j)], \quad (15)$$

where $[X(\cdot)]$, $[Y(\cdot)]$, and $[Z(\cdot)]$ represent rotations about the x, y and z axes, respectively. Using this convention, and the notation $\mathbf{d}^j=(d_{x,j}, d_{y,j}, d_{z,j})$, the position data can be expressed as the 4 x 4 homogeneous transform

$$[K(t)] = \begin{bmatrix} c\phi c\psi - s\phi s\theta s\psi & -c\psi s\phi s\theta - c\phi s\psi & c\theta s\phi & d_x \\ c\psi s\theta & c\theta c\psi & s\theta & d_y \\ c\psi s\phi - c\phi s\theta s\psi & -c\phi c\psi s\theta + s\phi s\psi & c\phi c\theta & d_z \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (16)$$

where $\sin(\cdot)$ and $\cos(\cdot)$ are noted with $s(\cdot)$ and $c(\cdot)$, respectively. Let the movement of the task frame M relative to the world frame F be defined by the 4 x 4 homogeneous transform $[K(t)]$, and consider its Taylor series expansion in the vicinity of both start and finish positions, such that

$$[K^j(t)] = [K_0^j] + [K_1^j]t + \frac{1}{2}[K_2^j]t^2 + \dots, \quad j = 1, \dots, n$$

where $[K_i^j] = \frac{d^i[K^j]}{dt^i} \Big|_{t=0}$. (17)

The matrices $[K_0^j]$, $[K_1^j]$ and $[K_2^j]$ are defined by the position, velocity and acceleration of the end-effector in the vicinity of the two task positions M^j . A point \mathbf{p} in the moving frame has the trajectory $\mathbf{P}^j(t)$ in the fixed frame F in the vicinity of a task position M^j (see Figure 1), given by the equation

$$\mathbf{P}^j(t) = [K^j(t)]\mathbf{p} = [K_0^j + K_1^j t + \frac{1}{2}K_2^j t^2 + \dots]\mathbf{p}. \quad (18)$$

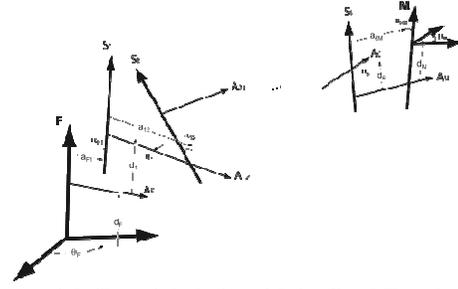


Figure 1. A spatial 6R serial chain with its fixed F and moving M frames

This equation can be rewritten by substituting $\mathbf{p}=[K_0^j]^{-1}\mathbf{P}^j$ to obtain the relative transformation

$$\begin{aligned} \mathbf{P}^j(t) &= [K_0^j + K_1^j t + \frac{1}{2}K_2^j t^2 + \dots][K_0^j]^{-1}\mathbf{P}^j, \\ &= [I + \Omega^j t + \frac{1}{2}\Lambda^j t^2 + \dots]\mathbf{P}^j. \end{aligned} \quad (19)$$

The kinematic specification consists of set of spatial displacements and the associated angular and linear velocities $[\mathbf{V}^j] = [\mathbf{W}^j, \mathbf{v}^j]$, $j=1, \dots, n$ in start and finish positions, where

$$[\mathbf{W}^j] = [\dot{A}^j][(A^j)^T], \quad \text{and} \quad \mathbf{v}^j = -[\mathbf{W}^j]\mathbf{d}^j + \dot{\mathbf{d}}^j, \quad j = 1, \dots, n. \quad (20)$$

The dot denotes derivatives with respect to time. From this, the 4 x 4 spatial velocity matrix $[\Omega^j]$ is given by

$$[\Omega^j] = \begin{bmatrix} 0 & -w_{z,j} & w_{y,j} & v_{x,j} \\ w_{z,j} & 0 & -w_{x,j} & v_{y,j} \\ -w_{y,j} & w_{x,j} & 0 & v_{z,j} \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad (21)$$

where $\mathbf{w}^j=(w_{x,j}, w_{y,j}, w_{z,j})$ is the angular velocity vector and $\mathbf{v}^j=(v_{x,j}, v_{y,j}, v_{z,j})$ is the linear velocity vector at the jth position. Assuming the acceleration properties of the motion are defined at the j-th position, yields to:

$$[\alpha^j] = [\dot{\mathbf{W}}^j], \quad \text{and} \quad \mathbf{a}^j = \dot{\mathbf{v}}^j. \quad (22)$$

In order to define the 4 x 4 acceleration matrix $[\Lambda^j]$, we introduce the 4 x 4 matrix constructed from (22), $[\dot{\Omega}^j] = [\alpha^j, \mathbf{a}^j]$, to obtain

$$[\Lambda^j] = [\dot{\Omega}^j] + [\Omega^j][\Omega^j], \quad (23)$$

where j denotes the position in which the acceleration terms are defined.

Spatial Synthesis Applications

The spatial synthesis example is a part of the authors' efforts to explore new, efficient methods for the design of fault-tolerant robot manipulators, as well as novel task-planning techniques. Particularly, the authors examined a non-redundant general six-degree-of-freedom TRS robot manipulator (see Figure 2), mounted on a movable platform is fault-tolerant with respect to the originally specified task, consisting of second-order specifications, after one of its joints fails and is locked in place.

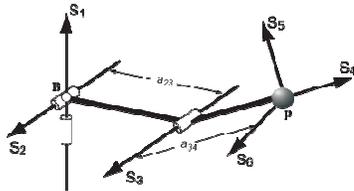


Figure 2. The TRS arm is a general six-degree of freedom serial chain configured so that the first pair of revolute joints intersect at right angles forming T-joint (also known as U-joint) and the last three joints intersect in a point to define a spherical wrist

The Denavit-Hartenberg parameters [20] of the arm are listed in Table 1. The task data is presented in Table 2. It consists of two positions with velocity, defined in the first position and velocity and acceleration specifications in the second position.

Table 1. Denavit-Hartenberg parameters for the TRS arm

Link	$\alpha_{i-1,i}$	$a_{i-1,i}$	θ_i	d_i
1	—	—	θ_1	0
2	$\pi/2$	650 mm	θ_2	0
3	0	0	θ_3	650 mm
4	$\pi/2$	0	θ_4	0
5	$\pi/2$	0	θ_5	0
6	$\pi/2$	0	θ_6	0

Table 2. Task data for planning movement of the TRS arm

Data	Start Position	Finish Position
Position (rad; mm)	$(0, -\pi/2, \pi; 0, 0, -350)$	$(\pi, \pi/2, 0; 0, -100, -800)$
Velocity (rad/s; mm/s)	$(0, 100, -300; -10, -100, -100)$	$(-100, 0, 0; -100, 10, 0)$
Accel. (rad/s ² ; mm/s ²)	—	$(0, 0, 0, -100, 10, 0)$

Figure 3 shows the TRS arm moving through the specified task. The trajectory is determined from the joint parameters using a fifth-degree polynomial interpolation following [21], [22].

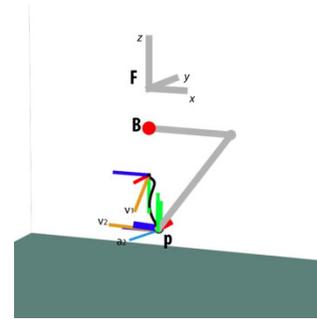


Figure 3. The world frame **F**, the location of fixed pivot **B** in the base of the arm, the moving pivot **p** of the TRS arm

In the following sections, the authors present the recovery strategy for the six-degree-of-freedom TRS arm, to achieve the originally specified task in the case of an actuator failure.

Arm Actuator Failures

The recovery strategy is based on the ability to reposition the arm base so that the point **B** at the intersection of S_1 and S_2 of the TRS can be placed where needed in a horizontal plane parallel to the X-Y plane of the world frame **F**. The point **B** lies at the origin of the fixed coordinate system, $B_z = -140$ mm at all times, i.e. the arm base can move freely in the X-Y plane. The authors assumed that the TRS arm could grasp the tool frame where necessary so that the wrist center **P** could be positioned in **F** where necessary. The proposed strategy reconfigures the arm-platform system using degrees of freedom that exist in the system but are locked during arm movement. Thus, the recovery plan is achieved by first identifying values for the reconfiguration parameters $\mathbf{B}=(B_x, B_y, -140)$ and $\mathbf{P}=(P_x, P_y, P_z)$ that ensure that the end-effector of the TRS arm can achieve the specified task for each particular arm joint failure, shown below. These constraint equations combine with the specified task to provide a set of polynomial equations for the reconfiguration parameters of the platform arm system. Solutions to these equations are obtained numerically using the polynomial homotopy continuation software PHC [23]. The movement of each reconfigured arm is determined by solving the inverse kinematics failure model in each of the task positions, and then using joint trajectory interpolation to guide its end-effector through the prescribed task [14].

Assume that the actuator of joint S_1 , in Figure 2, which controls the shoulder azimuth angle, of the TRS arm has failed and that the brakes have been set to maintain a constant angle θ_1 . The remaining actuated joints of the TRS form a parallel RRS chain that can position the wrist in a plane perpendicular to the horizontal axis. Once a normal $\mathbf{G}=(G_x, G_y, G_z)$ to this plane and a position of the wrist center $\mathbf{P}=(P_x, P_y, P_z)$ are identified, then the coordinates of the base pivot **B**

can be computed to reposition the base of the platform to allow the arm to guide the tool frame through the specified task, despite the S_1 joint failure [24]. The polynomial system of design equations for the parallel RRS consists of four bilinear quadratic equations and one linear equation in the five unknowns $\mathbf{r}=(G_x, G_y, P_x, P_y, P_z)$. The total degree of the system is $2^4 = 16$, which is small enough to directly eliminate the variables and obtain a univariate polynomial of degree six. The solution, corresponding to failure at $\theta_1=90^\circ$, is given in Table 3.

Table 3. The reconfiguration parameters for a failed S_1 joint

$\mathbf{B} = (B_x, B_y, B_z)$ (mm)	$\mathbf{P} = (P_x, P_y, P_z)$ (mm)
(0, 0, -140)	(0, -412.6, -487)

Figure 4(a) shows the reconfigured platform arm system and the trajectory generated to guide the TRS arm through the original task with an S_1 joint failure.

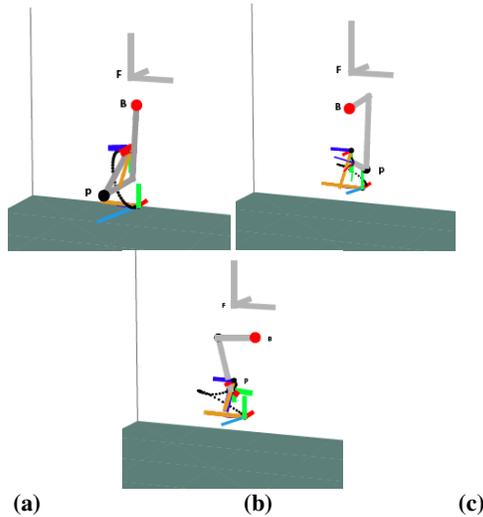


Figure 4. (a) The reconfigured platform arm system for an S_1 actuator failure. (b) The reconfigured platform arm system for an S_2 actuator failure. (c) The reconfigured platform arm system for an S_3 actuator failure

Next, the authors considered the case in which the actuator of the second joint S_2 of the TRS arm, which controls the shoulder elevation angle, fails and that the brakes have been set to maintain θ_2 at a constant value. The remaining joints of the arm in Figure 2 form a perpendicular RRS chain that can locate the wrist center on a circular torus. We obtain five polynomial equations that define the reconfiguration parameters $\mathbf{r}=(B_x, B_y, P_x, P_y, P_z)$ that allow the platform arm system to complete the task despite the failure. The system of five quartic polynomials has a total degree of $4^5 = 1024$. The real solution, corresponding to shoulder elevation failure at $\theta_2 = 0^\circ$, is listed in Table 4.

Table 4. The reconfiguration parameters for a failed S_2 joint

$\mathbf{B} = (B_x, B_y, B_z)$ (mm)	$\mathbf{P} = (P_x, P_y, P_z)$ (mm)	B_z^* (mm)
(62.31, -184.54, -140)	(21.19, 122.75, -466.16)	-140

Figure 4(b) shows the reconfigured rover arm system for the crippled TRS arm with an S_2 joint failure through the original task. If the elbow actuator of joint S_3 of the TRS arm in Figure 2 fails then we assume its brakes can be set so that θ_3 has a constant value. The remaining joints of the arm form a TS chain that can position the wrist center $\mathbf{p}=[K_{01}^T \mathbf{P}]$ on a sphere, with a radius R , about the base point \mathbf{B} . The radius R is defined by the link lengths a_{23} and a_{34} , the angle θ_3 , and is equal to

$$R^2 = a_{23}^2 + a_{34}^2 - 2a_{23}a_{34}\cos\theta_3, \quad (31)$$

where the value of θ_3 is determined from the joint sensor of the failed actuator. As in the previous cases, we seek the reconfiguration parameters $\mathbf{r}=(B_x, B_y, P_x, P_y, P_z)$ that allow the arm to perform the original task, despite the failure. The polynomial system consists of five quadratic equations in the unknowns \mathbf{r} and has a total degree of $2^5 = 32$ [25]. The real solution, corresponding to elbow failure at $\theta_3 = 68.56^\circ$, i.e. $R = 400$ mm, is given in Table 5.

Table 5. The reconfiguration parameters for a failed S_3 joint

$\mathbf{B} = (B_x, B_y, B_z)$ (mm)	$\mathbf{P} = (P_x, P_y, P_z)$ (mm)	R (mm)
(17.1, -124, -140)	(-33.4, 158, 418)	400

Figure 4(c) shows the reconfigured platform-arm system and the trajectory generated to guide the TRS with the elbow joint failure to achieve the originally specified task. A Surface Mobility Platform (Gears LLC), a Lynxmotion robot arm, integrated using Single-Board RIO – 9632 (National Instruments) are used for the experimental set up. Tests of the proposed strategy are currently performed in the Space Robotics Lab at Texas A&M (see Figure 5).

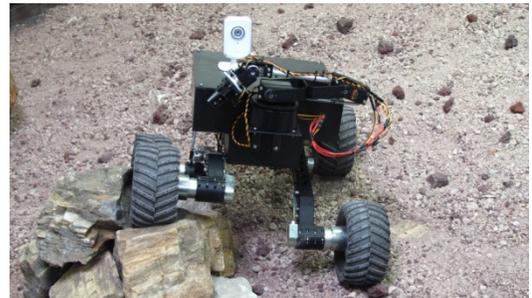


Figure 5. Experimental set up. Arm in stowed position

Figure 6 (a) shows the healthy arm holding a tool, moving through a task consisting of second-order specifications. Figure 6(b) shows the new location of the fixed \mathbf{B} and moving \mathbf{p} pivots of the arm, after an elbow failure. The closer

look shows how the re-grasping ability of the end-effector had allowed for the tool to be grasped at a different location.

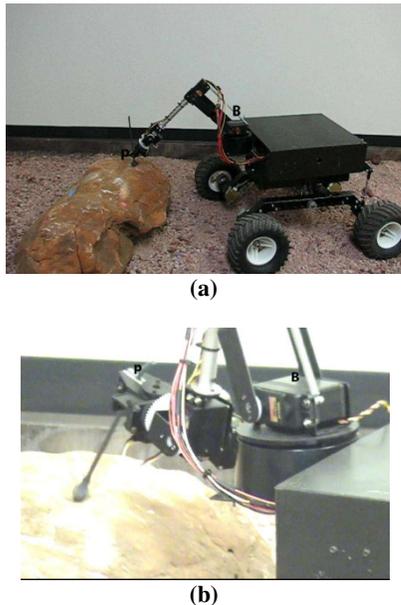


Figure 6. (a) The healthy arm moving through the a task. (b) New locations for the base pivot B and the moving pivot p have been obtained in order for the crippled TS arm to obtain the originally specified task despite the elbow failure

Finally, Figure 8 is a recent result from our efforts to design mechanical linkages, constrained to have the same coordinates for the fixed pivot **B** and the same end-effector path trajectory.

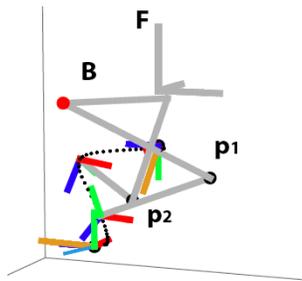


Figure 8. The TS chain (Bp1) and the perpendicular RRS chain (Bp2) move smoothly through the second order task

The animation shows that both TS and perpendicular RRS linkages satisfy the second-order task specification and move smoothly throughout the task.

Summary

Formulation of the kinematic specification for the synthesis of spatial kinematic chains with specified task acceleration was presented. Applications have focused on exploring

new strategies for the failure recovery of general six-degree-of-freedom manipulators. The last example combines the second-order effects of the task with the particular kinematics of the chain to yield free parameters that allow for more than one system to accomplish the same task.

References

- [1] D. Tesar and J. W. Sparks, "The Generalized Concept of Five Multiply Separated Positions in Coplanar Motion", *J. Mechanisms*, 1968, 3(1), 25-33.
- [2] H. J. Dowler, J. Duffy, and D. Tesar, "A Generalised Study of Four and Five Multiply Separated Positions in Spherical Kinematics—II", *Mech. Mach. Theory*, 1978, 13:409-435.
- [3] Chen P. and Roth B., "Design Equations for the Finitely and Infinitesimally Separated Position Synthesis of Binary Links and Combined Link Chains", *ASME Journal of Engineering for Industry*, 1969, Vol. 91: 209-219.
- [4] Hanafusa H., and Asada H., "Stable Prehension by a Robot Hand with Elastic Fingers", *Proc. 7-th Int. Symp. Ind. Robots*, 1977, pp. 384-389.
- [5] Cai, C.C., and Roth B., "On the Spatial Motion of a Rigid Body with a Point Contact", *IEEE Int. Conf. on Robotics and Automation*, 1987, pp. 686-695.
- [6] Montana, D.J., "The Kinematics of Contact and Grasp", *Int. Journal Robot. Research*, 1988, vol. 7, NO. 3, pp. 17-25.
- [7] Park, J., Chung W. and Youm Y., "Second Order Contact Kinematics for Regular Contacts," *IEEE Conference on Intelligent Robots and Systems*, 2005, pp.1723-1729.
- [8] Sarkar N., Yun S., and Kumar V., "Dynamic Control of 3D Rolling Contacts in Two Arm Manipulation", *IEEE Int. Conf. on Robotics and Automation*, 1993, pp. 978-983.
- [9] Trinkle J.C., "On the Stability and Instantaneous Velocity of Grasped Frictionless Objects", *IEEE Trans. Robotics and Automation*, 1992, vol.8, pp. 560-572.
- [10] Ohwovoriole M. S. and Roth B., "An Extension of Screw Theory", *J. Mech. Design*, 1981, Vol. 103, pp. 725-735.
- [11] Reuleaux F., *The Kinematics of Machinery*, 1963, New York: Dover.
- [12] Mishra B., Schwarz J.T., and Sharir M., "On the Existence and Synthesis of Multi-finger Positive Grips", *Algorithmica*, 1987, Vol. 2, pp. 541-558.
- [13] Markenskoff X., Ni L. and Papadimitriou C.H., "The Geometry of Grasping", *Int. J. Robot. Research*, 1990, vol. 9, pp. 61-74.

-
- [14] E. Rimon and Burdick J., "A Configuration Space Analysis of Bodies in Contact - I. 1-st Order Mobility", *Mechanism and Machine Theory*, 1995, Vol.30(6) : 897-912.
 - [15] E. Rimon and Burdick J., "A Configuration Space Analysis of Bodies in Contact - II. 2-nd Order Mobility", *Mechanism and Machine Theory*, 1995, Vol. 30(6): 913-928.
 - [16] N. Robson and McCarthy J. M., "Kinematic Synthesis with Contact Direction and Curvature Constraints on the Workpiece", *ASME Int. Design Engineering Conference*, 2007.
 - [17] N. Patarinsky Robson and J. M. McCarthy, "Synthesis of a Spatial SS Serial Chain for a Prescribed Acceleration Task", *Proc. IFToMM, The 12th World Congress on Mechanism and Machine Science*, June 18-21, 2007, Besancon, France.
 - [18] N. Patarinsky Robson, J. M. McCarthy and I. Turner, "Applications of the Geometric Design of Mechanical Linkages with Task Acceleration Specifications", *ASME Int. Design Eng. Tech. Conf.*, 2009, August 30-Sept.2, San Diego, CA.
 - [19] McCarthy, J.M., *Geometric Design of Linkages*, Springer-Verlag, 2000, New York.
 - [20] Hartenberg, R. S., and Denavit, J., *Kinematic Synthesis of Linkages*, 1964, McGraw-Hill, New York.
 - [21] Craig, J., *Introduction to Robotics, Mechanics and Control*, 1989, Addison - Wesley Publishing Co.
 - [22] N. Patarinsky Robson, J. M. McCarthy and I. Turner, 2009, "Exploring New Strategies for Failure Recovery of Crippled Robot Manipulators", *ASME/IFToMM Int. Conf. On Reconfigurable Mech. and Robots (ReMAR'09)*, June 22-24, London, UK.
 - [23] Verschelde, J., and Haegemans, A., "The GBQ-Algorithm for Constructing Start Systems of Homotopies for Polynomial Systems", *SIAM Soc. Ind. Appl. Math. J. Numer. Anal.*, 1993, 30(2), pp. 583 – 594.
 - [24] N. Patarinsky Robson, J. M. McCarthy and I. Turner, "Failure Recovery Planning for an Arm Mounted on an Exploratory Rover", 2009, *IEEE Transactions on Robotics*, 25(6), pp. 1448-1453.
 - [25] N. Patarinsky Robson, J. M. McCarthy and I. Turner, "The Algebraic Synthesis of a Spatial TS Chain for a Prescribed Acceleration Task", *Mechanism and Machine Theory*, 2008, 43: pp.1268-1280.

Biographies

NINA ROBSON is an assistant professor in Manufacturing and Mechanical Engineering Technology at Texas A&M University. Her research is in robotics, kinematics of motion, and biomechanics. Her e-mail address is ninarobson@tamu.edu.

J. M. MCCARTHY is a professor in Mechanical Engineering at University of California, Irvine. His research is in design of mechanical systems, computer aided design, and kinematic theory of spatial motion. He can be reached at jmmccart@uci.edu.

APPLICATION OF SIX SIGMA TO GEAR BOX MANUFACTURING

Kambiz Farahmand, North Dakota State University; Jose Victorino Marquez Grajales; Mohsen Hamidi, North Dakota State University

Abstract

This study applies Six Sigma to optimize the keyway cutting operation on a shaft of a gear box produced by Horton Automatics Company. Since the operation had not been optimized, many problems had been raised in the assembly of keys with shafts and a considerable number of scrap parts had been produced which deteriorated profitability of the company. The main problem was that the keyway width did not allow keys to be assembled into the shaft tightly. To solve the problem and optimize the process, DMAIC (Define, Measure, Analyze, Improve, and Control) methodology of Six Sigma was applied. In this paper, the steps of DMAIC to optimize the operation are presented and illustrated. In the Define phase, the problem was defined and the specific operation of the manufacturing process was recognized by the PFMEA (Process Failure Mode Effects Analysis) technique. In the Measure phase, the operation was studied and the process capability was measured. In the Analyze phase, the factors that affect the keyway width were identified by a cause-and-effect analysis. In the Improve phase, the best combination of the levels of the factors was determined by using the Design of Experiments (DOE) technique and the best combination was applied. Finally, in the Control phase, some recommendations were given in order to keep the process in good condition.

Introduction

In statistical terms, "reaching Six Sigma" means that the process or product performs with almost no defects, but the real message of Six Sigma goes beyond statistics [1]. Six Sigma is a philosophy of managing that focuses on eliminating defects through practices that emphasize understanding, measuring, and improving processes [2]. The focus of Six Sigma is reducing variability in key product/service quality characteristics to the level at which failure or defects are extremely unlikely [3]. The model of a Six Sigma process assumes that if the process is centered at the target and the nearest specification limit is six standard deviations from the mean, the process will operate at the 3.4 parts-per-million defect level [3]. Six Sigma was heavily inspired by six preceding decades of quality improvement methodologies such as quality control, TQM, and Zero Defects [4], [5]. Motorola first made Six Sigma popular in the 1980s, AlliedSignal embraced it in the early 1990s and then General Electric made it the most popular management philosophy in history [6].

Six Sigma efforts target three main areas of improving customer satisfaction, reducing cycle time, and reducing defects [1]. Improvements in these areas usually represent dramatic cost savings to businesses, as well as opportunities to retain customers, capture new markets, and build a reputation for top-performing products and services [1]. Unlike mindless cost-cutting programs which reduce value and quality, Six Sigma identifies and eliminates costs which provide no value to customers, or waste costs [7]. Basu and Wright [8] listed some real benefits from the adoption of Six Sigma: for example, in 1997 Citibank undertook a Six Sigma initiative and after just three years it was reported that defects had reduced by ten times; General Electric reported that \$300 million invested in 1997 in Six Sigma delivered between \$400 million and \$500 million in savings, with additional incremental margins of \$100 to \$200 million; and Wipro Corporation in India says that two years after starting in 1999, defects were reduced to such an extent as to realize a gain of eight times over the investment in Six Sigma.

Six Sigma employs a well-structured program methodology; namely, Define, Measure, Analyze, Improve and Control (DMAIC) or Define, Measure, Analyze, Design and Validate/Verify (DMADV) [9]. DMAIC is used for projects aimed at improving an existing business process and DMADV is used for projects aimed at creating new product or process designs [10]. The five steps of DMAIC are as follows [8]. Figure 1 shows a flow diagram of the steps of DMAIC.

- 1) Define opportunities: This is done through identifying, prioritizing, and selecting the right projects.
- 2) Measure performance of the projects and process parameters.
- 3) Analyze opportunities: Opportunities are analyzed by identifying key causes and process determinants.
- 4) Improve performance: This is achieved by changing the process so as to optimize performance.
- 5) Control performance: This is essential if gains are to be maintained.

In the following sections, the steps of DMAIC applied to the manufacturing process of a gear box produced by Horton Automatics Company are described. Since 1960, when they developed the first automatic sliding door in America, Horton Automatics Company has been designing and manufacturing automatic systems such as automatic sliding, swinging, folding, and security revolving doors, service windows,

presence and motion detection systems, Automated People Mover (APM) transit doors, and vehicle door operators.

is the C7113 subassembly, which includes a C7034 shaft and a C7017 gear. There were many failures occurring in the C7113 subassembly. The main problem was that the key did not fit in the keyway and this made a considerable number of scrap parts which had to be reworked or discarded.

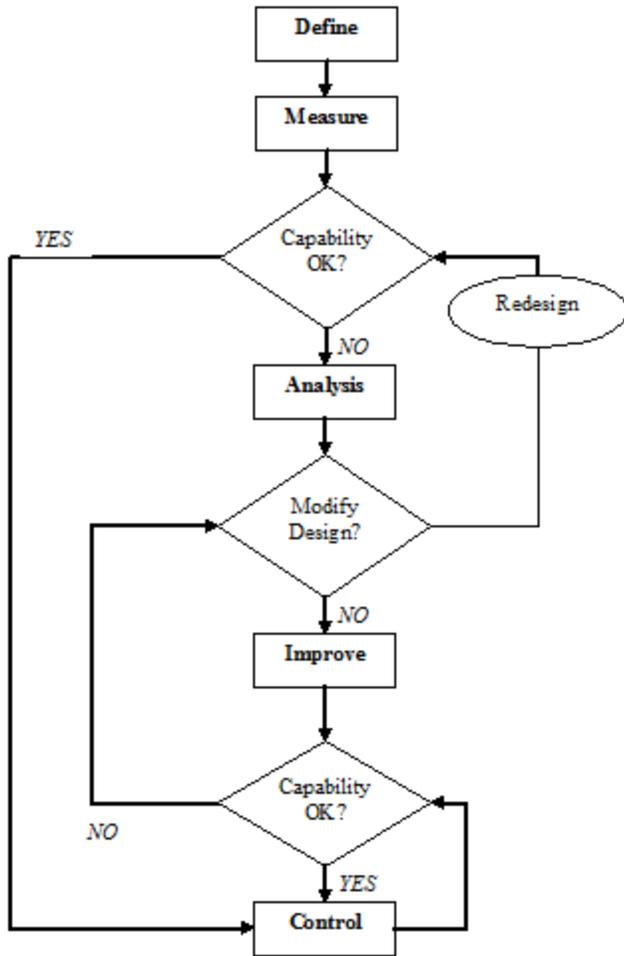


Figure 1. Flow Diagram of DMAIC

Define Phase

The automatic swinging door uses the 7000 gear box in order to open and close the door. The gear box includes three shafts and three gears. Each gear is attached to a shaft by the employment of a key. The most common type of keys is flat key, as shown in Figure 2. On the shaft, the keyway is classified according to the process by which it is made. Figure 3 shows three of the most common keyways. The main function of the key is to transmit torque from a component to a shaft (see Figure 4).

In one part of the production line, gears and shafts are manufactured and in the other part, gears, shafts, keys, and other components, such as casings and switches, are assembled together. One of the main subassemblies of the product

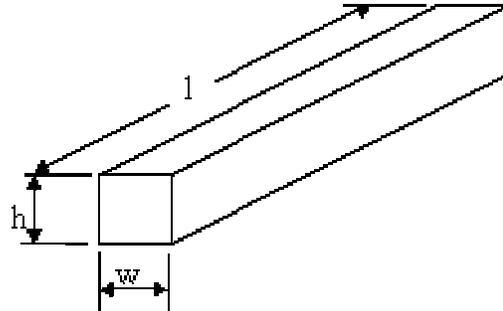


Figure 2. A Flat Key (w: Width, h: Height, l: Length) [11]

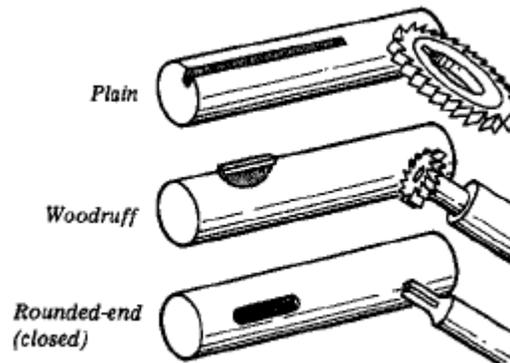


Figure 3. Three Types of Keyways [11]

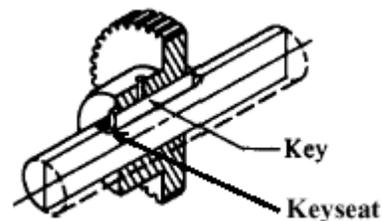


Figure 4. Keys Role for Transmitting Torque to a Shaft [11]

The primary purpose of the Define phase is to ensure that the team focuses on the right things [9]. In order to identify all of the problems in the manufacturing processes of the shaft, PFMEA (Process Failure Mode Effects Analysis) was applied. The main processes on the shaft are a hobbing process, performed by a hobbing machine, to cut teeth on the shaft and a cutting process, performed by a milling machine called “key machine”, to cut the keyway on the shaft. The PFMEA for these two processes is shown in Figure 5.

Through PFMEA, the following items were identified or determined for each process:

- Potential failure modes.
- Potential effects of failure and severity (SEV) rankings of the consequences of failure; the severity ranking is based on a relative scale ranging from 1 to 10 [12].
- Potential causes of failure and occurrence (OCC) rankings based on how frequently the cause of the failure is likely to occur; the occurrence ranking scale is based on a relative scale from 1 to 10 [12].
- Process controls for detecting the failure and detection (DET) rankings based on the chances the failure will be detected prior to the customer finding it; the detection ranking scale is on a relative scale from 1 to 10 [12].
- The overall risk of each failure which is called Risk Priority Number (RPN): $RPN = SEV \times OCC \times DET$; the RPN (ranging from 1 to 1000) is used to prioritize all potential failures to decide upon actions leading to risk reduction.
- Recommended actions to reduce the risk (RPN).

The PFMEA in Figure 5 indicates that the key machine process, which is performed by a milling machine, has the

highest RPN. Based on the observations in this study, it was found that the method of cutting the keyway on the shaft had not been studied and validated by the plant and there was much confusion in the way to cut the keyway on the shaft. A vertical milling machine was used to cut the keyway. The operator frequently changed the machine setup and generated different dimensions in the width of the keyway; consequently, many keys could not be assembled to the shafts tightly and properly due to incorrect size of the keyway.

Measure Phase

In this phase, a process capability analysis was performed. A sample of 30 pieces was taken and Minitab software was used to do the analysis. C_{pk} index is a process capability index. Equation (1) shows the formula used for calculating C_{pk} . Usually, C_{pk} should be above 1.33

$$C_{pk} = \min \left[\frac{USL - Mean}{3\sigma}, \frac{Mean - LSL}{3\sigma} \right] \quad (1)$$

Line No.	Process Function / Requirements	Potential Failure Mode	Potential Effect(s) of Failure	SEV	OCC	Potential Cause(s) / Mechanism(s) of Failure	OCC	Current Process Controls Prevention (P) Detection (D)	DET	RPN	Recommended Action (s)
15											
16	Hobbing machine	Incorrect angle	Interference between gears	8		Error Operator	4	Red Line machine	3	96	
17											
18		Run Out	Egg Shape	8		Arbor vibration	4	Dial indicator	9	288	
19											
20		sharpening to much > 1 inch	Broke the hobb	7		Do not check the width of the teeth	9	Visual	9	567	Implement a gage to measure the width of the teeth of the hobb.
21											
22		Sharpening at wrong angle	Interference between gears	8		Error Operator	4	Red line machine	3	96	
23											
24											
25		Incorrect handle	Hobb Broken	8		No trained	7	NONE	9	504	
26											
27											
28	key machine	4 flu	To loose (2 times) some curve (1 time)	8		Many flu take off too much material	10	Caliper	9	720	Identify which are the critical variables and put it in control
29											
30		2 flu	2 times ok 0.187	8		Don't take enough material	10	Caliper	9	720	Identify which are the critical variables and put it in control
31											
32		move dimension	deeply way	6		wear out	7	caliper	9	378	
33											
34		forget key way on the part	already welded	7		Error Operator	2	NONE	9	126	

Figure 5. PFMEA Chart

In Figure 6, the results of the analysis are shown. The width of the keyway should be 0.188 (-0.000, +0.002) inches; LSL (Lower Spec Limit) is 0.188 inches and USL (Upper Spec Limit) is 0.190 inches. C_{pk} of the process was 0.31, indicating that the process was in a bad condition. As seen in Figure 6, the sample mean was larger than the expected mean (0.189), which means that the keyway width was usually larger than the width that is ideal for assembling the key onto the shaft. Also, some widths were outside the spec range, less than LSL or more than USL, and as a result many keys could not be assembled onto the shafts tightly and properly due to the incorrect width of the keyway.

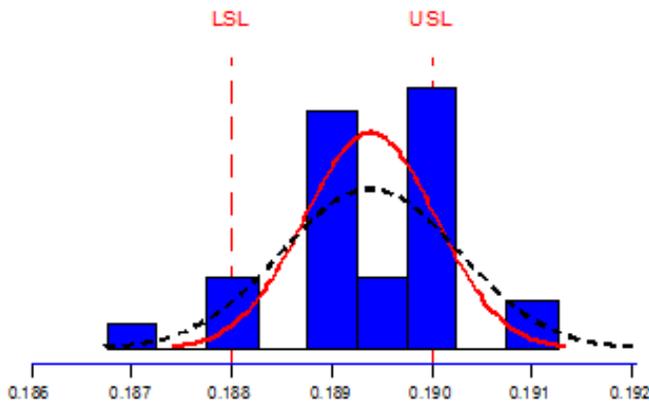


Figure 6. Process Capability Analysis

Analysis Phase

In order to find out the possible causes that affect the keyway cutting process, a fish-bone diagram was developed, as shown in Figure 7. The possible causes are divided in setup, operator, end milling, and machine. For setup, the variables to control are the center of cutter and the feed rate. The center of the cutter does not affect the keyway width, but the effect of the feed rate was unknown and the operator was confused about the suitable feed rate for the operation. The effect of the feed rate, then, had to be investigated. For operator, this is a factor that can be controlled with operator training. For end milling, there are three variables to control: wear out, the material of the cutter, and the number of flutes in the cutter. Wear out can cause imperfections in the keyway and there was no estimated time to change the cutter; it was only changed when it was broken. To control this cause, the supervisor was told to monitor the start date of a new cutter and how many pieces could be made with the same cutter. The types of cutters used were either carbide or high-speed steel, which are quite similar and do not affect the keyway width. Basically, the selection of the cutting tool (carbide or high-speed steel) was determined not to be a contributing factor to the incorrect dimension in the keyway. Also, there were two types of cutters: 2-flute cutters and 4-

flute cutters. The effect of the number of flutes was unknown and had to be investigated. For the machine factor, the variables to control are: the holder of the shaft, machine cleaning, and the head machine. The holder of the shaft should be locked but it is dependent on the operator. Cleaning is done by the operator but there was not a maintenance program. The head machine is an important factor but it is dependent on the operator; as long as it is locked and 90 degrees to the arm, it does not create any problems.

Therefore, the factors for which the effects were unknown were identified to be the number of flutes (2 or 4) in the cutter and the feed rate. The feed rate was determined by the RPM (Revolutions per Minute) of the milling machine. The rotational speed of the machine can be 1750 RPM or 2720 RPM. Basically, to improve the process, the effects of the number of flutes of the cutter and the speed of the machine had to be investigated.

Improve Phase

Design of Experiments (DOE) is a technique for examining controlled changes of input factors and the observation of resulting changes in outputs, i.e., the response to input changes [8]. DOE was applied in order to determine which factor(s) have major effects on the width of the keyway and what combination of factors' levels gives the best result. As mentioned earlier, there are two factors, number of flutes and RPM, each having two levels. Thus, the experimental design is called a factorial design. In statistics, a factorial experiment is an experiment whose design consists of two or more factors, each with discrete possible values or levels, and whose experimental units take on all possible combinations of these levels across all such factors. Such an experiment allows one to study the effect of each factor on the response variable, as well as the effects of interactions between factors on the response variable. For the vast majority of factorial experiments, each factor has only two levels. Table 1 shows the data gathered for the analysis. In each combination of the levels of the factors, there are ten replications.

The statistical hypothesis is as follows:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_1 : \mu_i \neq \mu_j ; \text{ For at least one pair of } i \text{ and } j (i, j=1,2,3,4)$$

where:

μ_1 : Width mean with 2 flutes and low RPM (1750 rpm)

μ_2 : Width mean with 4 flutes and low RPM (1750 rpm)

μ_3 : Width mean with 2 flutes and high RPM (2720 rpm)

μ_4 : Width mean with 4 flutes and high RPM (2720 rpm)

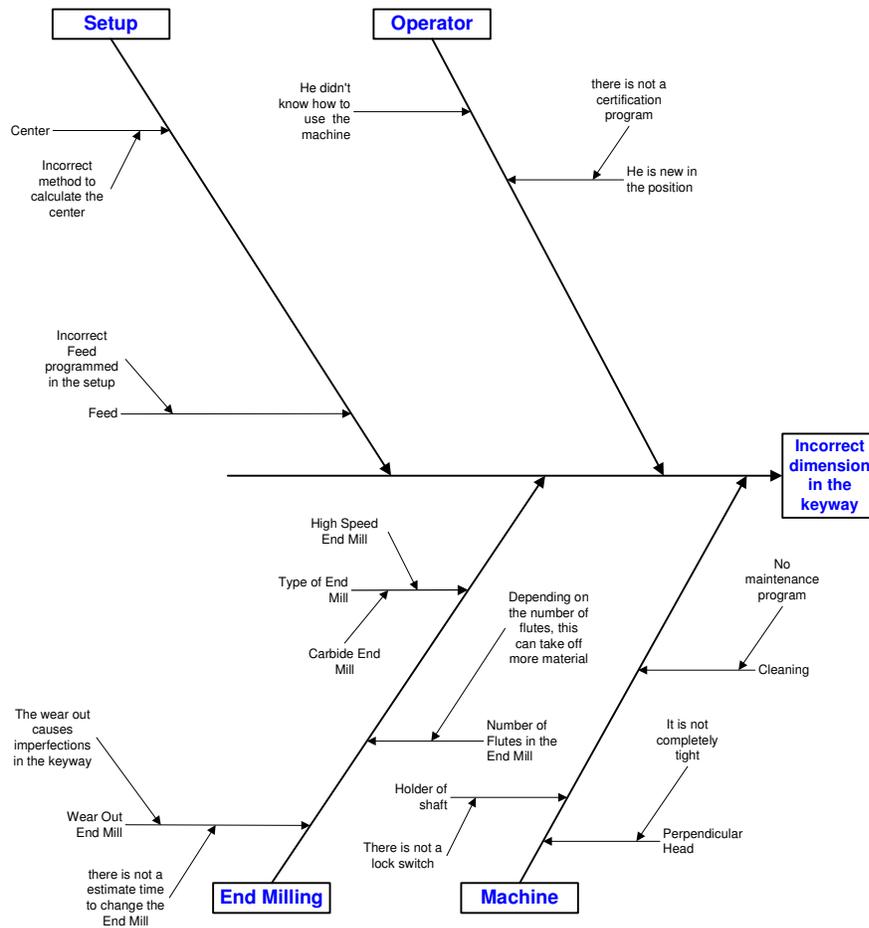


Figure 7. Fish Bone Diagram

Table 1. Width of the Keyway (in inches)

Number of Cutter Flutes (Factor A)	RPM (Factor B)	
	1750 (Level 1)	2720 (Level 2)
2 Flute (Level 1)	0.189, 0.188, 0.196, 0.187, 0.189, 0.188, 0.190, 0.190, 0.188, 0.190	0.190, 0.187, 0.187, 0.188, 0.186, 0.188, 0.188, 0.187, 0.188, 0.187, 0.187, 0.190
4 Flute (Level 2)	0.195, 0.195, 0.192, 0.197, 0.197, 0.196, 0.195, 0.196, 0.195, 0.194	0.194, 0.192, 0.194, 0.194, 0.193, 0.194, 0.193, 0.194, 0.193, 0.193

The descriptive results of the analysis performed by Mini-tab software is shown in Appendix 1. As seen in the first table of this appendix, the F test statistic for main effects (factor A: number of flutes and factor B: RPM) is 74.54 and the P value is 0.000. At a significance level of $\alpha = 0.05$, the null hypothesis (H_0) could be rejected because $P = 0.000 < \alpha = 0.05$ and $F = 74.54 > F_{0.05, 2, 36} = 2.84$. The result shows that there is a significant difference between the means of the four possible combinations of the factors' levels, and the effect of the main factors (Flute and RPM) on the keyway width. Since for the 2-way interaction the F statistic was very small and the P value was more than α , it was concluded that the factors do not have significant interactions. An interaction between factors occurs when the change in response from the low level to the high level of one factor is not the same as the change in response at the same two levels of a second factor. That is, the effect of one factor is dependent upon a second factor.

In the second table of Appendix 1, the individual effects of the factors were investigated. The P values for Flute and RPM were 0.000 and 0.001, respectively, and were both less than $\alpha = 0.05$. Therefore, it was concluded that both factors have significant effects on the keyway width. In Figure 8, the normal probability plot of the effects is shown. Effects (the related points) that do not fall near the line have a significant impact on the result. The Pareto chart in Figure 9 displays the magnitude and importance of the effects. Any effect that extends past the reference line (the dashed line) is important. As shown in Figures 8 and 9, both factors A and B have significant effects, but factor A (Flute) has the most significant effect. And the interaction effect of AB is not significant.

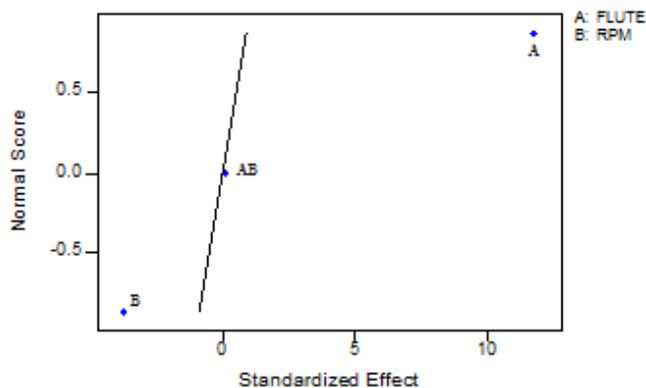


Figure 8. Normal Probability Plot of Effects

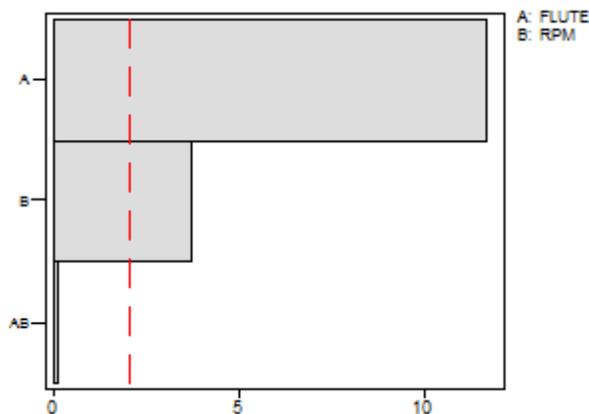


Figure 9. Pareto Chart of the Effects

Figure 10 shows the main effect plot. The main effect plot is a plot of the means at each level of a factor. The means for the factors' levels can be seen in the third table of Appendix 1. Since the interaction between Flute factor and RPM factor is not significant, the best combination of these factors can be found by using the main effect plot. As shown in Figure 10, the width increases significantly when the number of

flutes changes from 2 to 4. The mean for 2 flutes is 0.1884 and within the spec range (LSL = 0.188 and USL = 0.190), but the mean for 4 flutes is 0.1943 and outside the spec range. The mean for an RPM of 1750 is 0.1923, more than the USL and outside the spec range. But it decreases and falls in the spec range when the RPM is changed to 2720. In conclusion, the best setting for the operation is a 2-flute cutter with an RPM setting of 2720 for the machine.

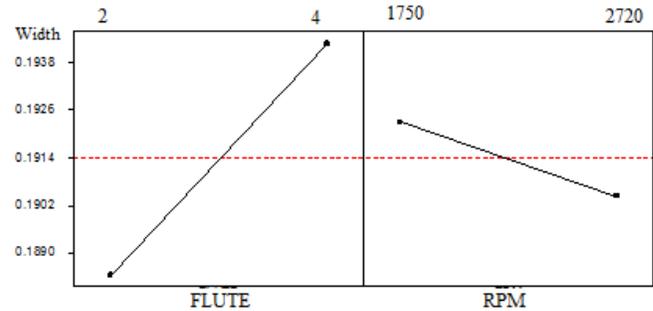


Figure 10. Main Effect Plot

In this phase, some parts were operated with the recommended levels of the factors (2-flute cutter, RPM of 2720) and the feed rate of 27.2 in/min. The feed rate was calculated by using equation (2):

$$\text{Feed Rate} = \text{RPM} * \text{Chip Load per Tooth} * \text{No. of Flute} = 2720 * 0.005 * 2 = 27.2 \text{ in/min} \quad (2)$$

For removing some metal particles inside the keyway and having accurate measurements, the edges of the keyway had been polished with sand paper. Since the polish operation on the shaft generated several burrs and spikes around the keyway, which caused the key not to fit within the keyway, a different polishing tool was recommended in order to eliminate all burrs within the keyway.

Control Phase

To control the dimension of the keyway, the operator should always use a 2-flute cutter and set the RPM of the machine to 2720. Also, it is necessary to have a good 2-flute cutter. If a used cutter is used, problems will be generated in the keyways. The supervisor should monitor the start date of a new cutter and how many pieces are made with the same cutter.

Conclusions

The problem was that the keyway width in the C7034 shaft of a C7113 subassembly in a 7000 gear box did not fit the width of keys used for its assembly. This problem led to a considerable number of scrap parts. DMAIC methodology

of Six Sigma was applied to solve the problem. In the Define phase, PFMEA was applied to distinguish the high-priority area of the process to focus on. The cutting operation performed by a vertical milling machine got the highest RPN and so the study focused on this area. In the Measure phase, the keyway cutting process on the shaft was evaluated and depicted by using a process capability measure of C_{pk} . As expected, the mean of the sample was larger than the anticipated mean (according to standard spec). In the Analyze phase, the cause-and-effect analysis was performed to recognize the main causes of the problem and the main factors that affect the keyway width and their levels as well. In the Improve phase, DOE was used to find the best combination of the levels of the factors. A factorial design analysis with two factors (number of flutes of the cutter and RPM of the milling machine) at two levels (2 and 4 flutes, 1750 and 2720 rpm) was performed. The results of the experiment provided by Minitab software showed that there was a sig-

nificant difference between the means of four possible combinations of the factors' levels. Both factors have significant effects and the most significant factor affecting the keyway width was the number of cutter flutes. The best combination of factor levels is having a 2-flute cutter and an RPM of 2720, which would provide a 27.2 in/min feed rate. Finally, in the Control phase, some recommendations were given in order to keep the process in good condition. The main recommendation is to trace the usage time of cutters because after a certain amount of time, a cutter cannot perform well. During this study, cooperation among engineers and operators and their excellent teamwork in the PFMEA analysis, process capability analysis, cause-and-effect analysis, and DOE were valuable. Actually, Six Sigma methodology produced a team that together would optimize the process and recognize and remove quality problems that had affected customer satisfaction and profitability of the company.

Appendix 1. Minitab Results of the Factorial Design

Analysis of Variance for Width

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	2	0.00037645	0.00037645	0.00018823	74.54	0.000
2-Way Interactions	1	0.00000003	0.00000003	0.00000003	0.01	0.921
Residual Error	36	0.00009090	0.00009090	0.00000253		
Pure Error	36	0.00009090	0.00009090	0.00000253		
Total	39	0.00046738				

Estimated Effects and Coefficients for Width

Term	Effect	Coef	SE Coef	T	P
Constant		0.191375	0.000251	761.70	0.000
FLUTE	0.005850	0.002925	0.000251	11.64	0.000
RPM	-0.001850	-0.000925	0.000251	-3.68	0.001
FLUTE*RPM	0.000050	0.000025	0.000251	0.10	0.921

Least Squares Means for Width

	Mean	SE Mean
FLUTE		
2	0.1884	0.000355
4	0.1943	0.000355
RPM		
1750	0.1923	0.000355
2720	0.1904	0.000355

References

- [1] Pande, P. S. & Holpp, L. (2002). *What is six sigma?*. McGraw-Hill.
- [2] Brue, G. (2002). *Six Sigma for managers*. New York: McGraw-Hill.
- [3] Montgomery, D. C. (2010). A modern framework for achieving enterprise excellence. *International Journal of Lean Six Sigma*, 1(1), 56-65.
- [4] Stamatis, D. H. (2004). *Six Sigma Fundamentals: A Complete Guide to the System, Methods, and Tools*. New York: Productivity Press.
- [5] Montgomery, D. C. (2009). *Statistical Quality Control: A Modern Introduction* (6 ed.). Hoboken, New Jersey: John Wiley & Sons.
- [6] Eckes, G. (2003). *Six sigma for everyone*. Hoboken, New Jersey: John Wiley & Sons.
- [7] Pyzdek, T. (2003). *The Six Sigma handbook: a complete guide for green belts, black belts, and managers at all levels*. New York: McGraw-Hill.
- [8] Basu, R., & Wright, J. N. (2003). *Quality beyond Six Sigma*. Boston: Butterworth-Heinemann.
- [9] Ray, S., & Das, P. (2010). Six Sigma project selection methodology. *International Journal of Lean Six Sigma*, 1(4), 293-309.
- [10] De Feo, J. A. and Barnard W. W. (2005). *JURAN Institute's Six Sigma Breakthrough and Beyond: Quality Performance Breakthrough Methods*. Tata McGraw-Hill Publishing Company Limited.
- [11] Technical drawings of Horton Automatics.
- [12] "10 Steps to Conduct a PFMEA". Resource Engineering, Inc.
http://www.qualitytrainingportal.com/resources/fmea/fmea_10step_pfmea.htm, retrieved December 3, 2010.

Biographies

KAMBIZ FARAHMAND is currently a professor and department head in the Industrial and Manufacturing Engineering department at North Dakota State University. Dr. Farahmand has more than 30 years of experience as an engineer, manager, and educator. He is a registered Professional Engineer in North Dakota and Texas and past president of IIE Coastal Bend Chapter. Dr. Farahmand may be reached at kambiz.farahmand@ndsu.edu

MOHSEN HAMIDI is currently a doctoral student and instructor in the Industrial and Manufacturing Engineering department at North Dakota State University. Mr. Hamidi may be reached at mohsen.hamidi@ndsu.edu

MICROSTEPPING CONTROL OF STEPPER MOTORS WITH DATA INTERPOLATION AND DIRECT VOLTAGE CONTROL

Dong-Hee Lee, Kyungshung University; Shiyong Lee, the Penn State University Berks Campus;
Tae-Hyun Won, Dongeui Institute of Technology

Abstract

This paper presents a novel approach for microstepping control of a low-power stepper motor (SM) for an automotive dashboard instrument application with a direct connection of a low-cost microprocessor. The microstepping operation of the SM is very important to indicating instrument applications, for example, meters for fuel, battery, speed, oil, and engine revolution. The proposed system uses pulse-width modulated output and digital output pins of a microprocessor for low-cost implementation. In order to perform a smooth positioning operation for an indicator application, a simple low-pass filter (LPF) and modified sine data table are introduced.

The modified sine data table is produced by injected harmonics to reduce the low-frequency harmonics of the SM phase voltage. Also proposed are the S-curve function for the smooth position reference and the amplitude control of the motor phase voltage depending on motor acceleration and deceleration speeds. The S-curve function provides a smooth response while maintaining high acceleration. The proposed voltage amplitude controller can change the motor phase voltage according to the motor speed to compensate the back-electromotive force (EMF) at high speed and to reduce the torque ripple in the low-speed region. The proposed system is implemented with a low-cost, 8-bit microprocessor without any external memory and power devices. The effectiveness of the proposed control scheme is empirically verified by a practical automotive dashboard instrument system.

Introduction

The SM is widely used in an open-loop position control system for its inherent stepping-position operation characteristics without any feedback loop [1-3]. Today, low-power and small-size SMs are extensively adapted as indicating instruments on the dashboards of automotive vehicles [4-6]. In an indicating instrument system, the SM controller should perform positioning smoothly. In order to achieve smooth positioning, a microstepping mode using a pulse-width modulation (PWM) approach is an excellent choice [3], [7-9]. Some applications use the multi-phase SM for smooth operation; however, the drive system of the multi-phase SM

is more complex than a conventional two-phase one [10-15]. In this paper, a low-cost dashboard indicating system is developed using a low-power SM which is directly connected to a microprocessor.

The proposed system uses the PWM output and the digital output pin of a microprocessor for each motor phase winding. The phase current can flow from the PWM pin to the digital pin during the positive voltage region, and it can flow in the opposite direction from the high active digital pin to the PWM pin in the negative voltage region.

In order to perform smooth positioning, a simple LPF and a modified sine data table are introduced to reduce the low-frequency harmonics of the motor phase voltage. The modified sine data table is produced by injected harmonics to reduce the harmonics of the motor phase voltage. Further proposed in this paper are the S-curve function for smooth position reference and amplitude control of phase voltage in connection with acceleration and deceleration. The S-curve function can generate the position reference with respect to the measured frequency for smooth indication with the SM.

The voltage amplitude controller can produce enough phase voltage to compensate for the back-EMF when the motor operates with high speed in the fast acceleration and deceleration region. Similarly, the amplitude of the phase voltage is decreased to reduce the torque ripple in the low acceleration and deceleration region from the low speed and low back-EMF of the motor.

The proposed system is designed with a low-cost, 8-bit microprocessor without any external memory and power device. The effectiveness of the proposed control scheme is verified with a practical automobile dashboard system. The experimental results show the smooth positioning of the dashboard indicator system.

Conventional Microstepping Operation

Figure 1(a) shows a conventional H-bridge circuit for an SM and its bipolar switching method for the microstepping operation. Different from the conventional full-step and half-step operation modes, the microstepping operation can indi-

cate the precise rotating angle, using sinusoidal phase voltage with PWM technology, as shown in Figure 1(b).

The output torque of the SM can be derived by the phase current and vector summation of each phase torque [7]. The phase current equation can be summarized as follows:

$$\begin{aligned} i_{as} &= I_m \cdot \cos \theta_e \\ i_{bs} &= I_m \cdot \sin \theta_e \end{aligned} \quad (1)$$

where, i_{as} and i_{bs} are phase currents of phase A and phase B. I_m denotes the current per phase. θ_e is the electrical position of an SM.

The resulting torque generated by the corresponding phases is derived by

$$\begin{aligned} T_{ma} &= K_T \cdot I_m \cdot \cos \theta_e \\ T_{mb} &= K_T \cdot I_m \cdot \sin \theta_e \end{aligned} \quad (2)$$

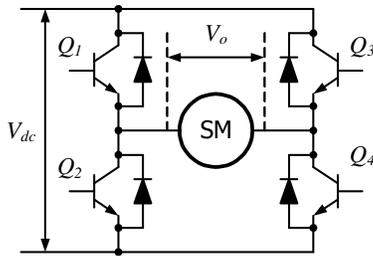
where K_T is the torque constant of the motor [7].

The total torque of the motor can be derived by the vector summation of the torque phases as follows:

$$(3) T_m = \sqrt{(T_{ma}^2 + T_{mb}^2)} = K_T \cdot I_m$$

In order to produce sinusoidal phase voltage and phase current, the conventional methods use a complex digital-to-analog converter (DAC) and a comparator circuit for PWM switching. The phase current is limited for stable operation.

In this paper, a simple microstepping operation scheme for the SM is directly connected to a low-cost microprocessor, and a passive LPF is introduced and applied to a dashboard indicator. The proposed scheme uses a simple modified sine table for reduction of low-frequency harmonics.



(a) H-bridge circuit per phase of a SM

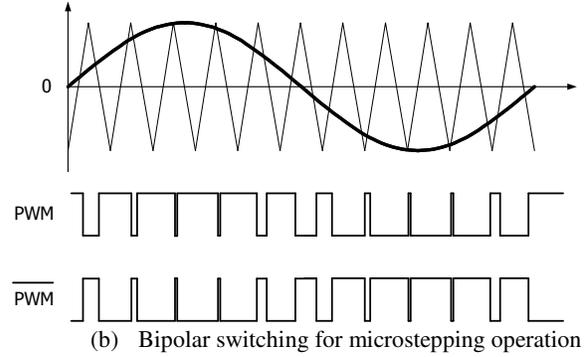


Figure 1. H-bridge circuit and bipolar switching for microstepping operation

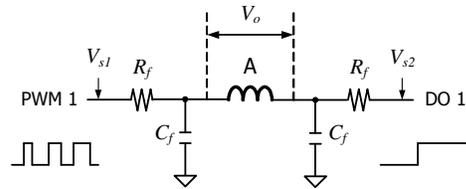
Proposed Direct Microstepping Operation (2)

PWM and Digital Output with a Passive Filter

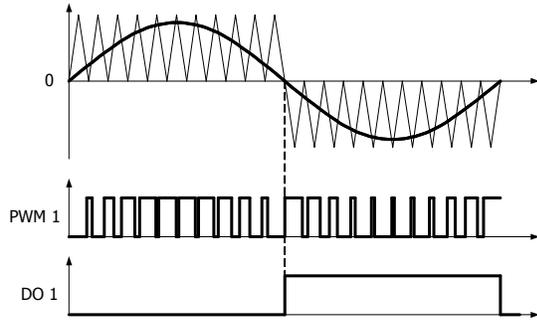
Also presented here is a direct microstepping operation scheme using PWM and the digital output of a low-cost microprocessor is proposed. In the front-end of the phase winding, a simple passive LPF is connected to reduce current ripple from PWM switching. The SM is designed for low-power consumption operating at 5V. The phase current is under 20mA, so the motor can be directly operated by the microprocessor output pin without an H-bridge converter or amplifier.

Figure 2(a) shows the passive LPF circuits for each phase and pin connection between the microprocessor and the SM. In Figure 2(b), 'PWM 1' and 'DO 1' are output pins of the microprocessor. In the positive voltage region, phase current can be generated by the 'Low' output of the digital pin and positive PWM output as shown in Figure 2(b). In the negative voltage region, phase current can be generated by the 'High' output of the digital pin and the negative PWM output.

The pulse duty ratio is controlled by the sine table according to the reference position of the SM and amplitude gain according to acceleration and deceleration. The digital output signal is changed according to the reference position. The digital output is kept at the 'Low' signal for 0 to 180 electrical degrees, and 'High' signal for 180 to 360 electrical degrees.



(a) Passive filter and pin connection per phase



(b) Signal output of PWM and digital output for sinusoidal waveform

Figure 2. The proposed phase connection and signal waveforms

The phase current flows from the PWM pin to the digital pin in the positive region, and flows from the digital pin to the PWM pin in the negative region. In the positive region from 0 to 180 electrical degrees, the terminal and phase voltages can be derived without the filter as follows:

$$\begin{aligned} V_{s1} &= d \cdot V_{cc} \\ V_{s2} &= 0 \\ V_o &= V_{s1} - V_{s2} = d \cdot V_{cc} \end{aligned} \quad (4)$$

where d is the duty ratio of the PWM and V_{cc} is the control voltage of the microprocessor.

In the negative region, from 180 to 360 electrical degrees, the terminal and phase voltages can be derived as follows:

$$\begin{aligned} V_{s1} &= (1-d) \cdot V_{cc} \\ V_{s2} &= V_{cc} \\ V_o &= V_{s1} - V_{s2} = -d \cdot V_{cc} \end{aligned} \quad (5)$$

From these relationships, the phase voltage can be controlled by the PWM duty ratio.

In order to supply a sinusoidal phase voltage, the internal memory of the microprocessor is used for the sine table. Figure 3(a) shows the output phase voltage and current with the passive LPF in conventional 8-bit PWM data.

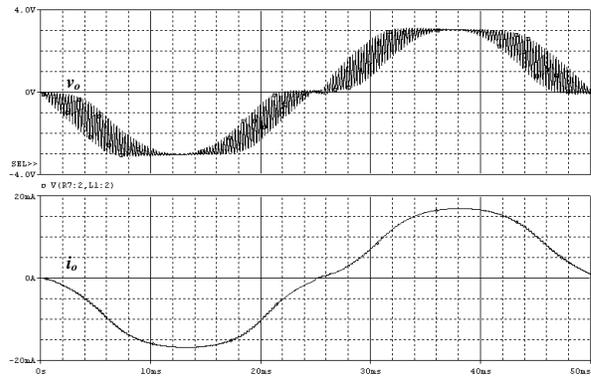
As shown in Figure 3(b), the conventional sine data produces low-frequency harmonics in phase voltage and current, such as 3rd and 5th harmonics. The switching harmonics

has a high-frequency component which slightly affects the position error. But, the low-frequency harmonics can produce an additional position error in the microstepping operation. In order to reduce the low-frequency harmonics, a modified sine table for 8-bit PWM signal is used. The modified sine table is generated with an additional harmonic injection to reduce the 3rd and 5th harmonic frequencies in the output voltage and current as follows:

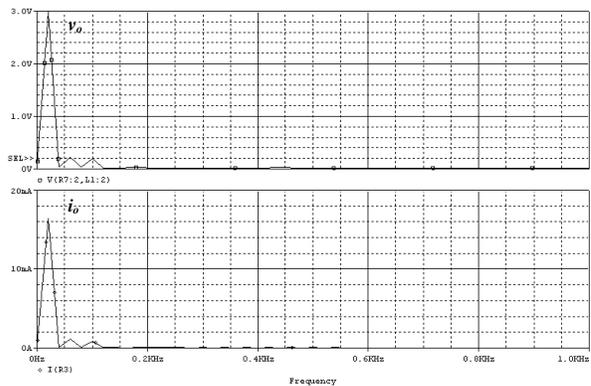
$$PWM[\theta] = 128 \cdot \sin(\theta) + a \cdot \sin(3\theta) + b \cdot \sin(5\theta) \quad (6)$$

where a and b are the injected harmonic coefficients of the 3rd and 5th harmonic frequencies, and where 6 and 4 are used for a and b respectively.

Figure 4(a) shows reduced low-frequency harmonics in the phase voltage and current with modified sine data. By comparison with Figure 3, the modified sine data with injected harmonics can reject the low-frequency harmonics. This pure sinusoidal phase current can produce a constant torque in any position.

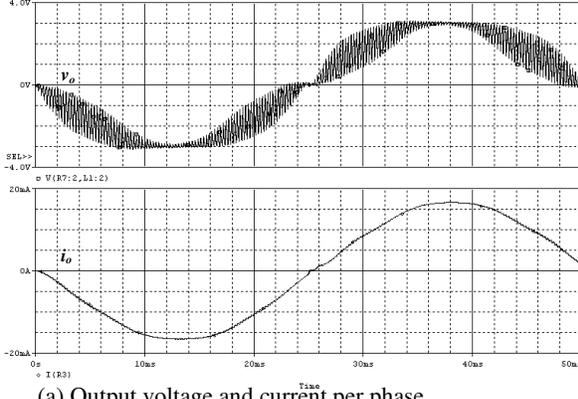


(a) Output voltage and current per phase

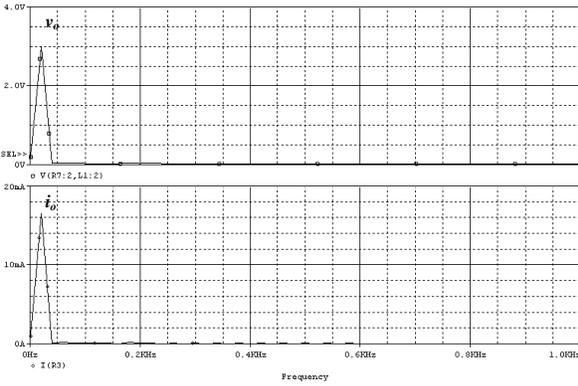


(b) FFT analysis of output voltage and current

Figure 3. Output characteristics of a conventional sine table



(a) Output voltage and current per phase



(b) FFT analysis of output voltage and current

Figure 4. Output characteristics of the modified sine table

Voltage Controller

Smooth torque control is essential in order to reduce the vibration of an indicator. In a conventional open-loop position controller of an SM, simple acceleration and deceleration curves are used without any current control. The conventional method is very simple, but the same torque at a different speed can cause indicator vibration.

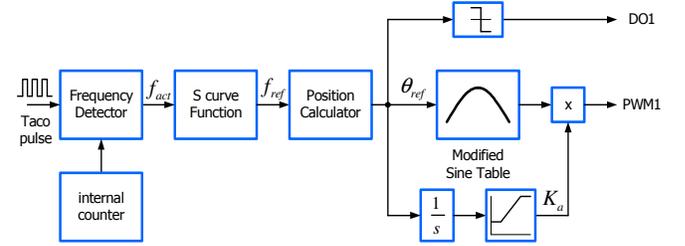
In this paper, a simple voltage control scheme to adjust the output torque of an SM according to motor speed is proposed. The variable voltage can change the phase current and output torque. The practical voltage controller is implemented by a PWM duty ratio control in order to change the PWM duty cycle by the position and motor speed and is given by

$$D = K_a \cdot \text{PWM}[\theta_{ref}] \quad (7)$$

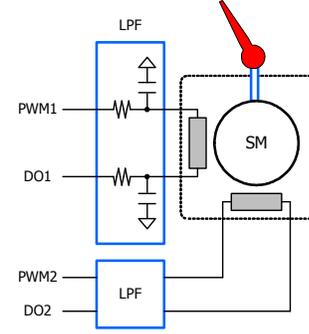
where D is the duty ratio of the PWM data. $\text{PWM}[\theta_{ref}]$ is the PWM data determined according to the reference position as described in equation (6). K_a is the proportional gain according to the motor speed and is lower than 1 as follows:

$$K_a = \lim(\omega_{ref} / \omega_{base}) \quad (8)$$

Figure 5(a) shows the proposed control scheme for the dashboard indicator. In order to reduce the indicating vibration, the phase voltage is controlled in relation to acceleration and deceleration. K_a , shown in Figure 5(b), denotes the amplitude gain of the PWM data according to the acceleration of the position reference. During fast acceleration, the amplitude gain is increased and the duty ratio of the PWM can be increased to increase the phase voltage.



(a) Proposed control scheme for a dashboard indicator



(b) SM connection

Figure 5. The proposed control scheme of a dashboard indicator using a low-power SM

For fast acceleration, the speed and the back-EMF of the SM should be increased with respect to the motor speed. In order to compensate the back-EMF in relation to increasing speed, phase voltage should be increased. In the proposed control scheme, the amplitude gain can control the phase voltage according to the motor speed. If the acceleration is decreased, the amplitude gain is decreased to decrease the phase voltage.

Experimental Results

In order to verify the proposed control scheme, an experimental test setup was implemented. A digital controller was designed using the ATmega16 8-bit microprocessor from ATMEL Corporation. The practical indicating instrument consists of one SM for indication and a 4-digit liquid-crystal display (LCD) for user display. The SM was directly con-

nected to the PWM output and digital output pins of the ATmega16 with a simple passive LPF for each phase. The SM has an internal gear train with a ratio of 180:1, and the mechanical accuracy is 0.3° per step. Figure 6 shows the implemented experimental test setup for the proposed control scheme.

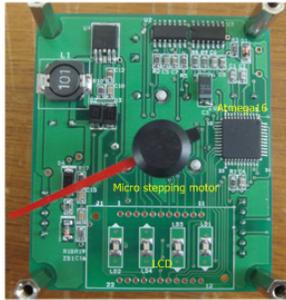
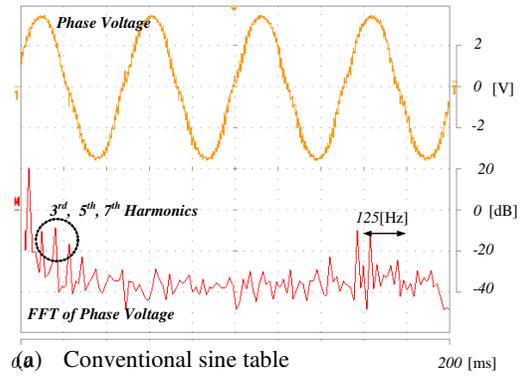


Figure 6. Experimental system

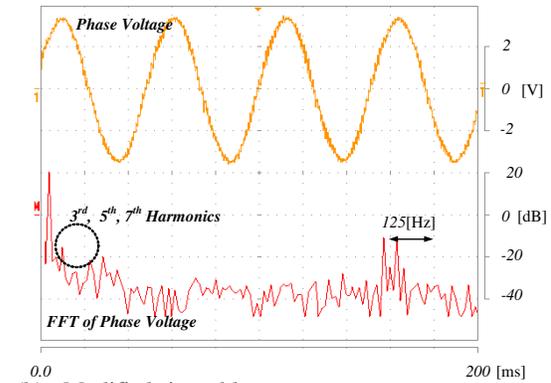
Figure 7 shows the phase voltage and the fast Fourier transformation (FFT) analysis of the SM under test with respect to the conventional sine and the modified sine data to reduce the low-frequency harmonics. As shown in Figure 7, the modified sine data can reduce the low-frequency harmonics with injected harmonics.

Figure 8 shows the experimental results at a 100Hz input frequency according to both sine tables. The measured frequency, the S-curve position reference, and the motor phase voltages are displayed in Figure 8. The amplitude of the phase voltage is fixed. The phase voltage waveform has low-frequency harmonics with the conventional sine table, as shown in Figure 8(a). However, the waveforms of the phase A and B voltages were closer to sinusoidal with the modified sine table without low-frequency harmonics, as shown in Figure 8(b).

Figure 9 shows the experimental results of the proposed control scheme at 100Hz and 300Hz input frequencies. The amplitudes of the phase voltages were changed by acceleration and deceleration. Compared with Figure 8(b) with a constant phase voltage, the amplitude of the phase voltage was not constant and the value was controlled by the speed, as shown in Figure 9. The low-phase voltage in the low-speed region reduces the phase current and the output torque. This is because the load torque of the indicator needle was almost the same, but the practical torque depends on the acceleration and deceleration torque. This variable phase voltage according to acceleration and deceleration can reduce the indicating vibration of the SM.

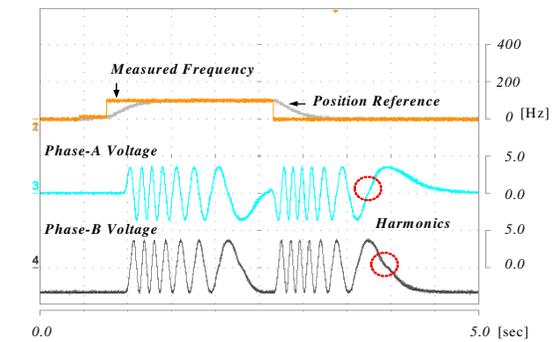


(a) Conventional sine table

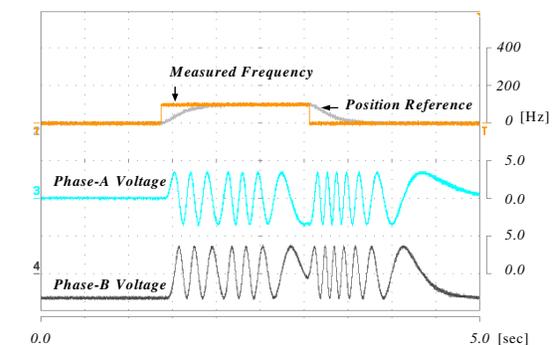


(b) Modified sine table

Figure 7. The phase voltage and FFT analysis



(a) Conventional sine table



(b) Modified sine table

Figure 8. Experimental results at 100Hz input

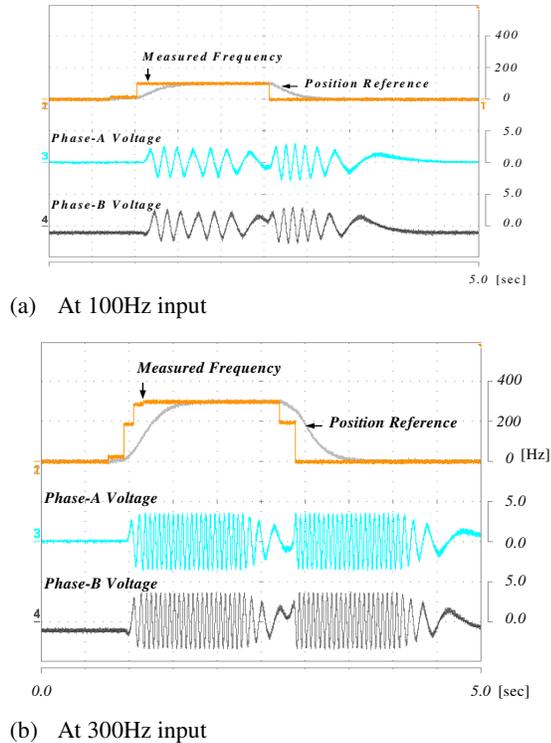


Figure 9. Experimental results of the proposed control scheme with two different input frequencies

Conclusions

This paper presents a microstepping operation of an SM which was connected directly to a microprocessor for an automotive dashboard indicator application. In order to implement smooth indicating, the modified sine data table with a simple LPF was used. The FFT analysis showed that the modified sine data can reduce the low-frequency harmonics significantly. In addition, an internal S-curve function for producing a smooth reference position from a detected input frequency was designed.

The amplitude of the phase voltage was controlled by the acceleration and deceleration of the reference position to reduce the indicating vibration. In the proposed control scheme, the amplitude of the phase voltage could be easily controlled by the amplitude gain, and the gain changes the actual duty ratio of the PWM data from the sine table.

The proposed system uses a PWM and digital output pins for each phase, allowing any low-cost microprocessor to be used for the dashboard indicator application. From the experimental results, the proposed system does a good job of controlling the SM with regard to the input frequency that eliminates low-frequency vibration.

Acknowledgment

The authors would like to express deep appreciation to Kyungsoo University, Busan, Korea, for its support in completing this study.

References

- [1] P. Acarnley, "Stepping Motors: a Guide to Modern Theory and Practice," 4th ed., IEE Control Engineering Series 63, ISBN: 0-85296-029-8, Michael Faraday House, 2002.
- [2] T. Kenjo and A. Sugawara, "Stepping Motors and Microprocessor Control," 2nd ed., ISBN: 0-19-859385-6, Oxford Clarendon Press, 2003.
- [3] A. Bellini, C. Concari, G. Franceschini and A. Toscani, "Mixed-Mode PWM for High-performance Stepping Motors," IEEE Transactions on Industrial Electronics, Vol. 54, No. 6, pp. 3167-3177, Dec. 2007.
- [4] Gildas Allain, "Micro Stepping: A Smoother Way of Life – Low-Cost Solution for Dashboard Instrumentation," Atmel Corporation, October 2008.
- [5] Data Instrumentation Technology, Ltd., "VID29 Series Cluster Stepper Motor Datasheets," www.vid.wellgain.com
- [6] MCR Motor, "Miniature Stepper Motor MR1107/MR1108 Datasheets," www.mcrmotor.com
- [7] A. Astarloa, U. Bidarte and A. Zuloaga, "Reconfigurable Microstepping Control of SMs using FPGA embedded RAM," in Proc. IEEE IECON, Nov. 2003, Vol. 3, pp. 2221-2226, 2003.
- [8] G. Baluta, "Microstepping Mode for SM Control," Signals, Circuits and Systems, 2007. ISSCS 2007. International Symposium, Vol. 2, pp. 1-4, July 2007.
- [9] G. Baluta, M. Coteata, "Precision microstepping system for bipolar SM control," Electrical Machines and Power Electronics, ACEMP '07. International Aegean Conference, pp. 291-296, Sept. 2007.
- [10] Keun-Ho Rew and Kyung-Soo Kim, "A Closed-Form Solution to Asymmetric Motion Profile Allowing Acceleration Manipulation," IEEE Transaction on Industrial Electronics, Vol. 57, no. 7, pp. 2499-2506, July 2010.
- [11] Tzung-Cheng Chen and Yung-Chun Su, "High Performance Algorithm Realization on FPGA for SM Controller," SICE Annual Conference 2008, pp. 1390-1395, August 2008.
- [12] Renesas R8C/25 Application Note, "180 Degree Sinusoidal Motor Control," REU05B0087-0100/Rev.1.00, pp. 1-13, February 2009.
- [13] K. S. Chun, H. J. Kim, Y. K. Kwon and S. H. Kang, "Design of High Performance 5 Phase Step Motor

-
- Drive System with Current Control Loop,” The Transactions of the Korean Institute of Power Electronics ,Vol. 11 no.5 , pp. 447-453, 2006.
- [14] M. H. Kim, T. Y. Kim, H. K. Ahn and S. K. Park, “One-Chip and Control System Design of Low Cost for Micro-stepping Drive of 5-Phase Stepping Motor,” The Transactions of the Korean Institute of Power Electronics , Vol. 9 no.1, pp.88-95, 2004.
- [15] S. H. Kim, E. W. Lee, D. J. Lee and T. M. Koo, “Microstep Drive of 2 Phase 8 Pole HB Type Linear Pulse Motor for Precise Position Control,” The transactions of the Korean Institute of Electrical Engineers. B, Vol. 48, no.12 , pp. 671-678, 1999.

Biographies

DONG-HEE LEE was born on Nov. 11, 1970 and received B.S, M.S., and Ph. D degrees in Electrical Engineering from Busan National University, Busan, Korea, in 1996, 1998 and 2001, respectively. He worked as a Senior Researcher of Servo R&D Team at OTIS-LG, from 2002 to 2005. He has been with Kyungsung University, Pusan, Korea, as an Assistant professor in the Department of Mechatronics Engineering since 2005. His major research field is the Power Electronics and motor control system. Dr. Lee may be reached at leedh@ks.ac.kr.

SHIYOUNG LEE is currently an Assistant Professor of Electrical Engineering Technology at The Pennsylvania State University Berks Campus, Reading, PA. He received his B.S. and M.S. degrees in Electrical Engineering from Inha University, Korea, his M.E.E.E. in Electrical Engineering from the Stevens Tech., Hoboken, NJ, and his Ph.D. degree in Electrical and Computer Engineering from the Virginia Tech., Blacksburg, VA. He teaches courses in Programmable Logic Controls, Electro-Mechanical Project Design, Linear Electronics, and Electric Circuits. His research interest is digital control of motor drives and power converters. He is a senior member of IEEE, as well as a member of ASEE, ATMAE, and IJAC. He may be reached at sul28@psu.edu.

TAE-HYUN WON is a professor of Electrical Engineering at Dongeui Institute of Technology, Busan, Korea. He received his B.S., M.S., and Ph.D. degrees in Electrical Engineering from Busan National University, Korea. He teaches Automatic Control, Electronic Circuits, and Micro-processor courses. His research interests include the sensorless control of PMSM, robust control of BLDC motor, and high precision and high speed motor control. He is a senior member of KIEE and KIPE. Professor Won may be reached at thwon@dit.ac.kr.

AN ENHANCED FRAMEWORK FOR MULTI-MODULE EMBEDDED RECONFIGURABLE SYSTEMS

Muhammad Z. Hasan, Texas A & M University; Timothy Davis, Texas A & M University; Troy Kensinger, Texas A & M University; Sotirios G. Ziaavras, New Jersey Institute of Technology

Abstract

Reconfigurable logic facilitates dynamic adaptation of hardware and ensures better utilization of hardware space as desired in embedded applications. Partial reconfiguration of hardware is a recent trend where a portion of the reprogrammable logic can be altered without affecting other portions. Host-based multiple-module reconfigurable hardware fabric, such as Field Programmable Gate Arrays (FPGAs), can potentially employ partial reconfiguration for embedded applications where a FPGA-resident or external host controls the application execution and reconfiguration. Although this technique minimizes area requirements and potential energy requirements for applications, it may result in a disparity in usage of different reconfigurable modules.

This disparity may cause localized temperature build-up and failure. Moreover, in such a host-based system, a subjective load distribution between the host and the reconfigurable module could result in performance improvement through parallelism. In this paper, policies are presented that ensure uniform utilization of reconfigurable modules, while implementing load-balancing between the host and the reconfigurable module for better performance. Experimental results involving benchmark kernels with these policies show a reduction in disparity of more than 40% of module usage as well as improvements in an application execution time of about 35%, as compared to a reference algorithm. In general, though, these policies are minimal when compared with the execution time for applications.

Introduction

FPGAs contain user-programmable hardware and interconnections. Thus, the reprogrammable features of FPGAs make it easy to test, debug, and fine tune hardware designs for higher performance in follow-up versions. Also, it enables the hardware implementation of a large design in a piecewise fashion as the complete design may not fit in the system. Partial reconfiguration support of current FPGA architectures provides support for reconfiguring portions of the hardware while the remainder is still in operation [1], [2]. Switching configurations between implementations can

then be fast, as the partial reconfiguration bit-stream may be smaller than the entire device configuration bit-stream.

Embedded systems are currently in virtually all aspects of everyday life. They normally are expected to consume small amounts of power and to occupy few resources. Numerous embedded applications spend substantial time on a few software kernels [3]. Executing these kernels on customized hardware could reduce the execution time and energy consumption as compared to software realizations [4], [5]. Given reconfigurable hardware, such as FPGAs, a chosen area could accommodate such kernels exclusively at different times to conserve resources, thus saving space and possibly power. Configurations to support kernels can be created ahead of time and stored in a database for future use, facilitating system adaptability for run-time events. However, the reconfiguration time affects the performance, especially for small execution data sets. Also, the reconfiguration process draws power. To offset the overhead time encountered, various techniques such as configuration pre-fetching or overlapping reconfiguration with other tasks must be employed.

Many dynamically reconfigurable systems involve a host processor mainly for control-oriented, less computation-intensive tasks and also for supporting reconfiguration decisions [4], [6-9]. The target of this work was either a single FPGA embedded with reconfigurable modules or several individually reconfigurable FPGAs. It was also shown that in a system with multiple reconfigurable resources, the disparity of usage may be significant under a brute-force policy [10]. In order to overcome such an undesirable effect, the runtime system could keep statistics for the utilization of each reconfigurable unit, in either a partially reconfigurable module or a complete FPGA. The ones with lower utilization should be the target for the next kernel implementation. This would not only balance the usage of all the reconfigurable resources but could also reduce the localization of temperature increases in the system for enhanced reliability.

Also, it was implied that when kernel execution on hardware, or the host, provides speedup, the whole data set for that kernel would be processed either on the hardware or on the host [10]. Instead of processing the whole data set for a kernel solely on the host or on the reconfigurable resources, it might be worth splitting the workload between them. This

type of load distribution and parallel execution might further boost application performance.

In this paper, the issues of uniform utilization of reconfigurable resources in a host-based multiple-module system and the load balancing between the host and the reconfigurable resource are considered. In Section 2, the scope of the problem and the proposed policies are defined. Explanation of the experimental set up including the simulation environment used to evaluate different policies is presented in Section 3. Results of this study are given in Section 4. Section 5 draws the conclusions.

Scope of the Problem

In this study, host-based dynamically varying embedded systems that change behavior at run-time and/or process time-varying work-loads were considered. The target of this study was either a single FPGA embedded with reconfigurable modules or several individually reconfigurable FPGAs. Such a framework [10] considered reconfiguration overheads in making decisions for the execution of kernels, either on the host or the FPGA(s), thereby ensuring performance gains. Considering the overhead of reconfiguration, the FPGA execution of kernels may not always be favorable, especially for small data sets. Thus, it addresses the issue of selective FPGA execution of kernels and reports performance improvement for synthetically generated kernel-based applications.

In a host-driven multiple-module reconfigurable system, the available reconfigurable hardware resources cannot often accommodate simultaneously all of the application kernels. As such, switching among kernels realized in hardware is necessary in real time. This may create non-uniform usage among different reconfigurable modules as data sets and the execution time of various kernels may differ. In order to selectively implement application kernels and to appropriately replace kernels, a methodology was proposed [10], [11]. The methodology resulted in improved application execution time. The core Break-Even (BE) policy of this methodology contains the following steps for a given kernel; these steps are repeated until all of the kernels of the application (program graph) are scheduled:

1. Estimate the execution time t_H on the host of the ready-to-execute kernel.
2. Check if the present FPGA configuration is the one required by the kernel. If 'yes', then set $t_{\text{overhead}} = 0$ and go to the next step.
3. If $t_H \leq t_{\text{overhead}} + t_{\text{comm}} + t_{\text{FPGA}}$, then execute the kernel on the host and exit. Else, proceed to the next step.

4. Reconfigure, if $t_{\text{overhead}} \neq 0$, an appropriate FPGA with the customized kernel configuration.
5. Transfer any necessary data from the host to the FPGA for execution.
6. Upload the results from the FPGA.

To place a ready-to-execute kernel in an appropriate FPGA, this methodology follows these steps:

1. Check if any FPGA is completely available. If 'yes', then place the kernel in this FPGA and exit. Else, proceed to the next step.
2. For each FPGA, compare the present kernels with the tasks/kernels in a window containing a preset number of kernels following the current kernel in the task graph. If there is a match, proceed to the next FPGA to repeat this process. Else, implement the kernel on this FPGA.

The above methodology is expected to enhance application execution performance as compared to host-only or FPGA-only execution. However, no specific policy is presented in this methodology to uniformly utilize the available reconfigurable resources. Moreover, the kernel execution could be expedited by parallelizing it between the host and the reconfigurable hardware. However, a befitting policy has to be in place in order to minimize idle time of the working entities. This issue has not been considered, although the issue of application execution performance is addressed to some extent [10], [11].

The original BE (Break-Even) policy presented by Hasan and Ziavras in a previous study was extended with two variations in order to achieve lower peak disparity of module usage [10], [11]. They are as follows.

Uniform utilization I

In this policy, uniform utilization of its resources was given preference over performance improvement of application. An FPGA, or a module, was chosen with the objective of ensuring uniform utilization. Candidate modules are the ones having lower utilization. With the highest utilized module, all of the candidate modules produced disparity greater than a threshold. The first module within this group was chosen for next-kernel execution. Then, a decision was made on executing a task on the host or on the reconfigurable logic considering the associated overhead.

Uniform utilization II

In this policy, performance improvement of application was given preference over uniform utilization of its resources. This policy differs from the above in two ways. First, the

FPGA, or a module, was chosen from the candidate module that had minimum utilization at that point. This would ensure uniform utilization. If this choice did not provide better performance than the host, then a second choice was made by considering a reconfigurable module that may hold that particular kernel. This would reduce reconfiguration overhead. Thus, although this policy targets uniform utilization, it may compromise that in favor of better performance.

Moreover, as mentioned earlier, an appropriate distribution of a data set between the host and the reconfigurable hardware could further boost application performance. In this respect, the data set is split for each kernel of an application such that the complete execution time on the host and on the reconfigurable logic (including any overhead) is virtually equal. However, under this methodology there are two options for choosing a reconfigurable module for hardware execution of the factored data set. A hardware module may be chosen to ensure uniform utilization among all. This choice ensures better uniformity of usage. This policy is called *load balancing with uniform utilization* and was not considered in the authors' previous studies [10], [11]. Alternatively, one may look ahead in the task graph to find the next two upcoming kernels for execution. A hardware module is chosen for execution that does not contain those upcoming kernels. This ensures reduced reconfiguration overhead and improved performance. This policy is called *load balancing with look-ahead*. These policies are implemented and evaluated in an embedded reconfigurable environment and results are presented in the sections to follow.

Experimental Set-up

The platform used to implement the embedded kernels, and to test the authors' current methodology, was the Starbridge Systems HC-62 Hypercomputer [12]. This system is a programmable, high-performance, scalable, and reconfigurable computer. It consists of eleven Virtex II FPGAs, of which ten are user programmable. In conjunction with the host, the HC-62 uses FPGAs to process complex algorithms. Although it does not mimic an embedded system, it can be used to simulate such an environment. Each FPGA can be thought of as an individual, partially reconfigurable module. VHDL designs can be imported into this environment by creating appropriate EDIF net list files. Xilinx tools are used to create configuration bit streams for the FPGAs. These bit files can be used to program them using a utility. The host can communicate with the FPGAs using appropriate PCI interface hardware and a second utility.

Application profiling of various EEMBC benchmarks resulted in kernel identification [13]. Due to earlier work on vector processing for embedded applications, focus is on such kernels [1]. They are: Autocorrelation between two

vectors, RGB-to-YIQ conversion, and High Pass Grey Filtering (HPG). MiBench [14] is a similar suite from the University of Michigan. The kernels chosen for experimental use from this suite were: 2D-DCT shuffling and FFT reordering [11].

The above five kernels were implemented on the hardware of the HC-62 system and their functionality was tested. Various data sizes were considered for each kernel, emulating dynamic load during execution. Each kernel behavior was also coded in C/C++ and the code was executed in advance on the host processor. Discussion of the execution times for these kernels on an HC-62 FPGA (XC2V6000), and on the host Xeon processor operating at 2.6GHz and having 1GB of RAM, can be found in a previous study by Hasan and Zia-vas [10].

The communication time between the host and the reconfigurable module was calculated considering the data volume to be transferred, the PCI bus interface clock frequency (66/133MHz), and the bus size (64-bits). The resulting values were quite accurate for the HC-62 host-FPGA system and can be used to validate the policy with experimental data. The reconfiguration time for the FPGA was 162ms.

Application cases were considered that solely consisted of computation-intensive kernels. Many application test cases were first created by randomly generating task graphs composed of the above five kernels. A publicly available program called Task Graphs For Free (TGFF) was used to generate these graphs [15]. The generated task graphs have many forks. These synthetically generated application task graphs were run on the host and the FPGA following the policies proposed by the authors in this current study. A sample application task graph composed of five different kernels is shown in Figure 1. Here, each node represents an application kernel and the arrows show the dependence among various kernels in the application. The size of an application task graph is the total number of kernels present in the graph. A range of sizes were considered from small (16 kernels) to very large (249 kernels).

As the actual setup of the HC-62 system unfortunately did not support partial reconfiguration, kernel implementations were considered on individual FPGAs of the HC-62 system to evaluate the proposed policies. A simulator was developed that takes the task graph, actual execution times of the host and FPGA, and the overheads of reconfiguration and data communication, as inputs. It mimics various execution behaviors of the system and calculates the respective execution times, number of reconfigurations, amount of disparity in usage, and the performance improvement for various policies. The developed simulation environment is shown in Figure 2.

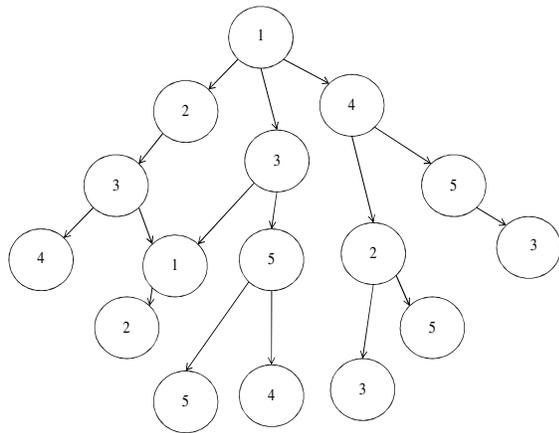


Figure 1. A Sample Application Task Graph (numbers represent kernel type)

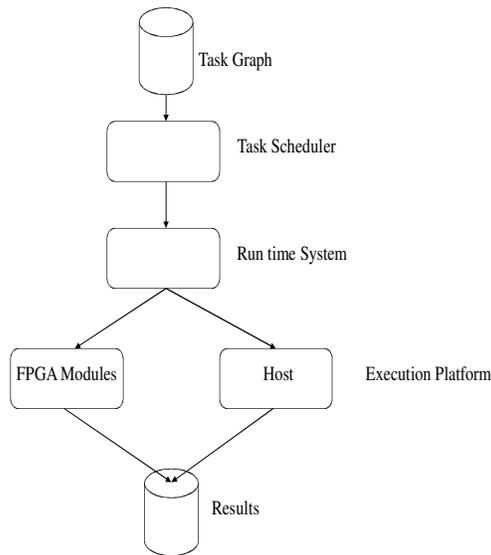


Figure 2. Simulation Environment

The task scheduler generates a correct sequence of execution for task-kernels appearing in the graph. It takes into account the dependencies among them to prepare the correct execution sequence. A kernel can be executed on the host as software, or on the FPGA as hardware. The choice is finalized at runtime by estimating the overall execution time of that kernel for the known data size without ignoring the overheads.

The run-time system decides whether to execute a kernel on the host or on the FPGA in order to ensure uniform utilization or performance improvement following a particular policy. It also chooses the appropriate FPGA or a module that should be reconfigured, when necessary. The simulator

was developed from scratch using the Visual C/C++ programming environment. It did not use any precompiled library routines from any other sources. This simulator runs on a PC having 1.67GHz Intel Core 2 Duo processor, 4GB of RAM, and operating under Windows Vista x64, Service Pack 2.

Results and Analysis

The two uniform utilization policies, as described in section 2, were simulated in the above environment and the results are summarized in Figure 3. Both of the variations of the original policy resulted in lower disparity in reconfigurable module usage, as indicated by the graph in Figure 3. For the largest application task graph of size 249, the reduction in disparity was significant, i.e., greater than 40%, as compared to the original BE policy. However, the peak disparities under the two extended policies are reasonably comparable except for one or two cases, also shown in Figure 3.

The execution time of each application task graph on the reconfigurable fabric was also determined from the simulation environment. These are plotted against the task graph sizes under different policies in Figure 4. A close look at this figure reveals that the execution times are almost inseparable from each other for the three policies in discussion. Although the original BE policy does provide lower execution time for a couple of task graphs, the resulting savings in execution time may not justify the higher peak disparity of usage it produces for reconfigurable resources. As such, it may be concluded that both the proposed policies targeted towards uniform utilization of resources would be equally desirable in such an application environment. However, if uniform utilization is a priority, the first policy would be preferred.

The experiment was also extended with a variable number of reconfigurable resources to observe the effect on the application execution time. Four execution kernels were considered to form application task graphs, and reconfigurable resources were varied between two to five. The observed trend is plotted in Figure 5. As can be seen from this graph, the execution time requirements fall off with increasing resources as would normally be expected. However, this trend holds only up to the point where the number of kernels in the task graph is equal to the available reconfigurable resources. If the reconfigurable resources are more than the execution kernels, the execution time remains unaffected and the curve becomes flat.

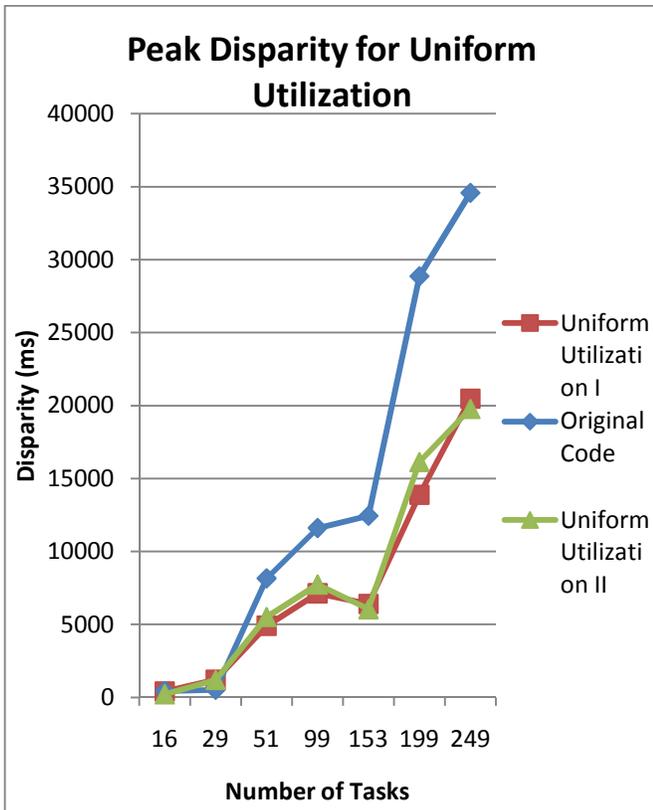


Figure 3. Peak Disparity under Three Policies for Uniform Utilization

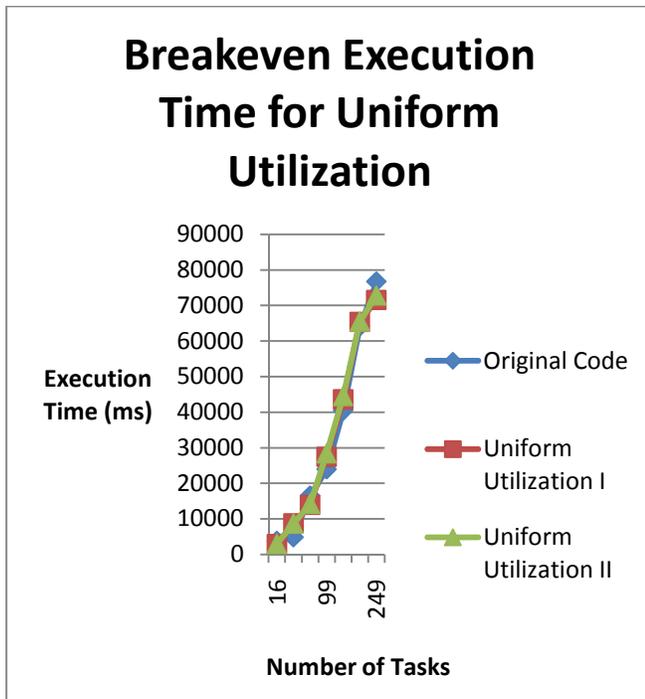


Figure 4. Execution Time under Three Policies for Uniform Utilization

Actually, in Figure 5, a slight upward trend can be seen beyond four reconfigurable resources. This is due to the policy of ensuring uniform utilization that tends to use an empty resource to reduce disparity of usage. As such, an unnecessary reconfiguration overhead might be encountered, thereby extending the overall execution time.

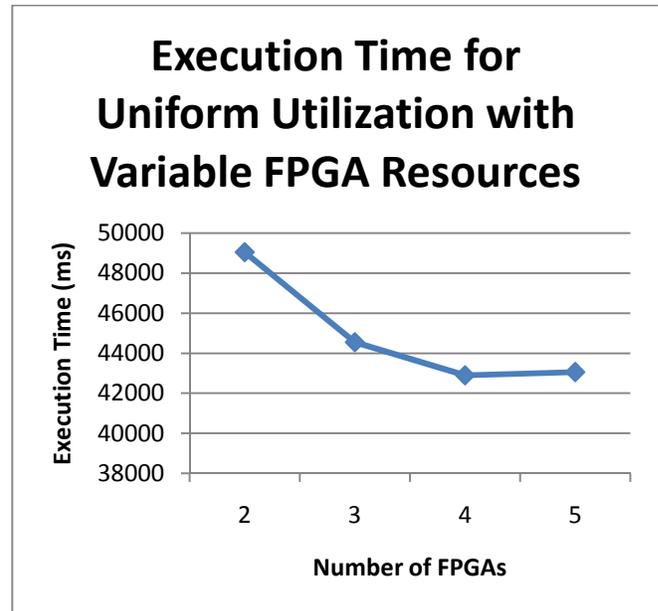


Figure 5. Execution Time of a Task Graph with Variable Reconfigurable Resources

Both of the variations of load balancing policies described in section 2 were then simulated within the current test bench. Let x be the total amount of data and y ($y < x$) be the amount processed by the FPGA hardware. If the rate of data processing on the host and on the FPGA are R_{host} and R_{FPGA} , respectively, then: $(y - x) / R_{\text{host}} = t_{\text{overhead}} + t_{\text{comm}} + x / R_{\text{FPGA}}$. This equation was used to find the value of y in order to equalize the execution time on the host and on the FPGA. The results of the simulation of load-balancing policies are given in Figures 6 and 7. Figure 6 shows the execution time for different application task graphs under these policies. As expected, the load-balancing strategy provides performance gain, especially for large task graphs. The execution time savings are significant (a drop from 7600ms to 5000ms or about 35%) for task sizes of 249 under load balancing with a look-ahead policy. However, the performance gain is not that prominent under load balancing with uniform utilization, as seen in Figure 6.

The peak disparity of reconfigurable resource usage was also evaluated under these two extended policies. As seen in Figure 7, peak disparity is quite low for load balancing with uniform utilization. However, as this policy suffers from performance degradation when compared to load balancing with look-ahead, the choice should be made among the two

policies based on the application needs. If performance gain is desirable, then load balancing with the look-ahead policy stands out. Otherwise, load balancing with uniform utilization should be chosen when minimization of disparity is desired.

So, for the vast majority of the tasks, both the host and the reconfigurable logic finish execution of their share of the load almost simultaneously. As such, the idle time is minimized. This reveals the reason for good performance gain under this policy.

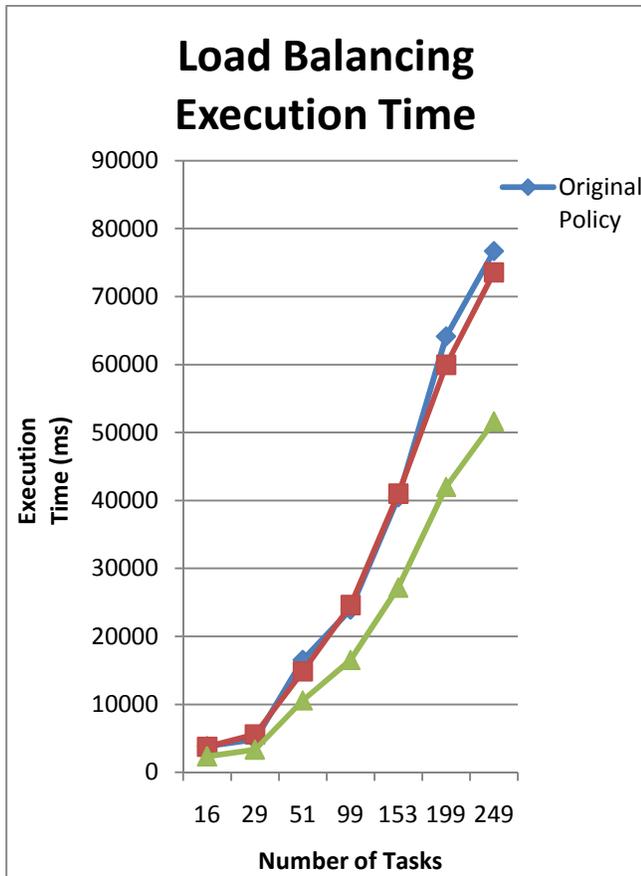


Figure 6. Execution Time under Two Policies for Load-balancing

In order to ensure good load balancing, it is of paramount importance that the host and the reconfigurable logic complete processing their share of data at about the same time. If they finish far apart in time from each other, then the total execution time will include the idle time of the one entity that completes first. This would certainly degrade the overall performance. However, the execution time of the respective payloads on the host and the reconfigurable logic should be comparable, including any overhead. This is illustrated in Figure 8, where the measured difference in execution time between the two entities for various tasks in an application under load balancing with the look-ahead policy is shown. As can be seen, for 84% (43+41) of the tasks, this difference is less than 5ms. Only for 3% of the tasks is the difference above 50ms.

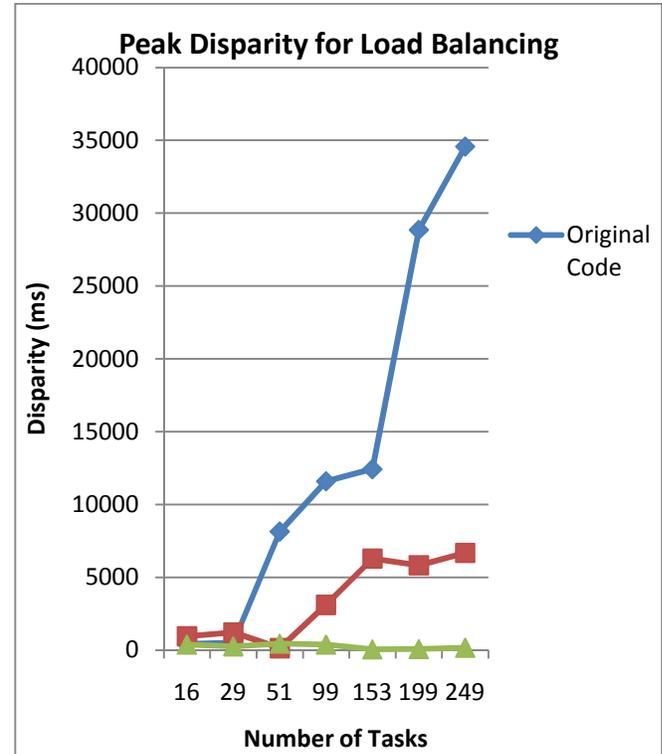


Figure 7. Peak Disparity under Two Policies for Load-balancing

There is an inherent difference in execution speed between the host and the reconfigurable hardware. In order to achieve true parallelism, then, the amount of data processed by the two entities within a given time is different. The total payload needs to be unevenly factored among them considering the overhead of communication and reconfiguration (if any). This is evident from Figure 9, where the distribution of the payload between the host and the reconfigurable logic for each task of an application task graph of size 29 is shown. Although the payloads of the two entities for task number 5 and for task number 8 are same, they are different for all of the other 27 tasks. In some cases, this difference is quite significant.

The processing time for different application task graphs were also measured using the time-function provided in the C/C++ developer's environment. This time includes, among others, the time to read the task graph from a file, scheduling the tasks considering their dependence, calculating the execution time under a selected policy, and writing the results back into a file.

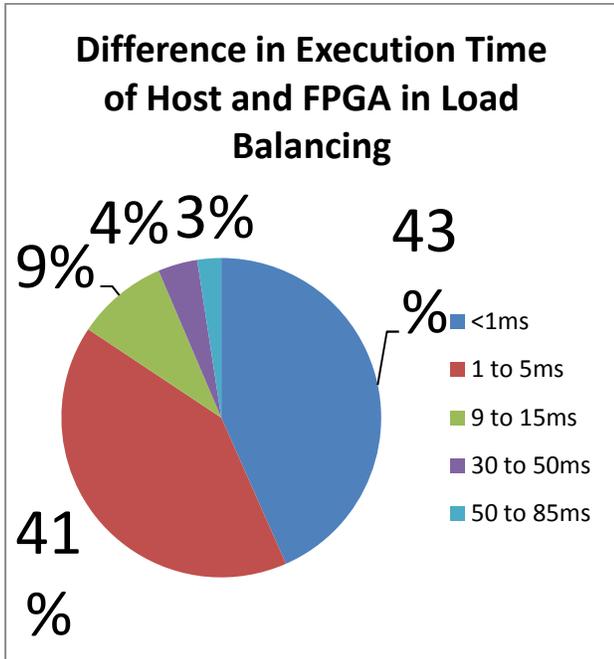


Figure 8. Differences in Execution Time under Load-balancing

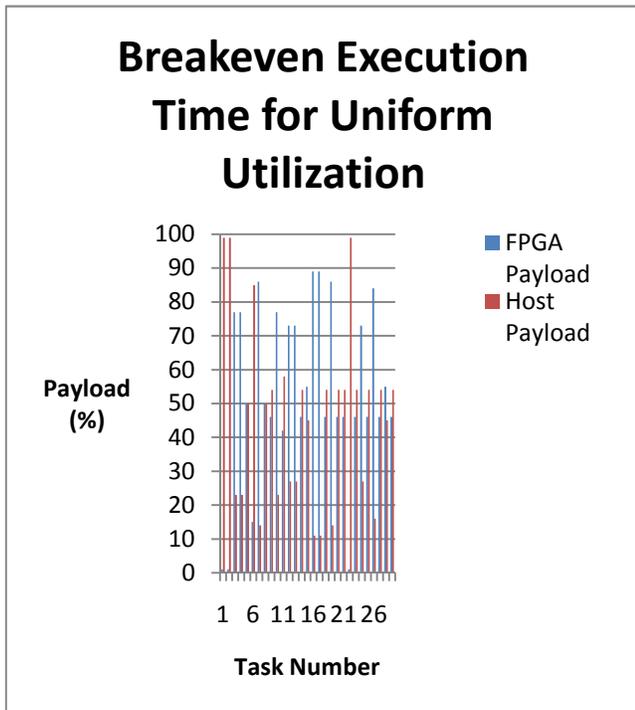


Figure 9. Payload Distribution among the Host and the Reconfigurable Logic

In a real host-based reconfigurable system, the last two components would not be present. So, the actual time needed by the host for all of the overhead would be less than the measured processing time. However, the measured

processing time was compared with that of the task graph execution time of Figure 6 in order to get a pessimistic feel for the host-overhead. Host-overheads were 2, 2.33, 3.67, 5, 7, 7.67, and 9ms for 16, 29, 51, 99, 152, 199, and 249 task graphs, respectively. A plot of the ratio of these two times is portrayed in Figure 10.

It clearly demonstrates that this host-overhead is contained within 9% of the application execution time even for very small task graphs. For the large task graph of 249, this overhead was less than 2%. It was believed that this figure would be even smaller for dedicated embedded systems. This is because the machine that was used for processing was a general-purpose one running many operating system tasks/threads in the background. As such, the time measurement of any program execution would most likely include the effect of others. So, an embedded host system executing a single program would be free from such side effects and, as a result, would imply much lower overhead than was measured in these current experiments.

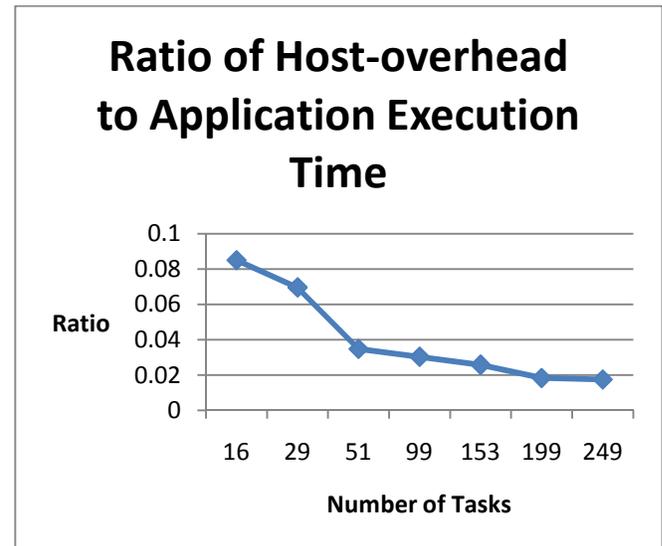


Figure 10. Ratio of Host-overhead to Application Task-graph Execution Time

Conclusions

Host-based reconfigurable systems have many benefits that are suitable for the embedded domain. Such systems were considered with multiple-reconfigurable modules executing kernel-based applications. The focus was on two objectives: obtaining uniform utilization among all reconfigurable resources, and achieving higher performance by proper load balancing between the host and the reconfigurable resources. Two policies were formulated to meet each of the above objectives. Simulation results revealed that the uniform utilization policies had comparable effects on reducing peak disparity and execution time. Both reduced the

peak disparity by about 40% as compared to the original reference policy. Also, one of the load-balancing policies, look-ahead, had superior effects over the other one with a reduction in execution time of about 35%. Furthermore, the overhead introduced for enforcing these policies was about 1% of the application execution time and, as such, can be incorporated into the embedded domain. Future research could include the investigation of using multiple reconfigurable resources to execute a single kernel and also to consider energy consumption under these circumstances.

References

- [1] M.Z. Hasan and S.G. Ziavras, "Runtime Partial Reconfiguration for Embedded Vector Processors," *Intern. Conf. Information Technology New Generations*, Las Vegas, Nevada, April 2-4, 2007.
- [2] M. French, E. Anderson, D. Kang, "Autonomous System on a Chip Adaptation through Partial Runtime Reconfiguration," *IEEE Symp. Field-Programmable Custom Computing Machines*, Palo Alto, California, April 14-15, 2008.
- [3] R. Krashinsky, et al., "The Vector-Thread Architecture", *31st Intern. Symp. Computer Architecture.*, Munich, Germany, June 19-23, 2004.
- [4] I. Robertson and J. Irvine, "A Design Flow for Partially Reconfigurable Hardware", *ACM Trans. Embedded Computing Systems*, 3(2), 2004, 257-283.
- [5] F. Barat, et al., "Reconfigurable Instruction Set Processors from a Hardware/Software Perspective", *IEEE Trans. Software Engineering*, 28(9), 2002, 847-862.
- [6] R. Lysecky, G. Stitt, and F. Vahid, "Warp Processors", *ACM Transactions on Design Automation of Electronic Systems*, 11(3), 2006, 659-681.
- [7] S. Ghiasi, et al., "An Optimal Algorithm for Minimizing Run-time Reconfiguration Delay", *ACM Trans. Embedded Computing Systems*, 3(2), 2004, 237-256.
- [8] W. Fu and K. Compton, "An Execution Environment for Reconfigurable Computing", *13th IEEE Symp. Field-Programmable Custom Computing Machines*, Napa, California, April 17-20, 2005.
- [9] B. Greskamp, and R. Sass, "A Virtual Machine for Merit Based Run-time Reconfiguration", *13th IEEE Symp. Field-Programmable Custom Computing Machines*, Napa, California, April 17-20, 2005.
- [10] M.Z. Hasan and S.G. Ziavras, "Customized Kernel Execution on Reconfigurable Hardware for Embedded Applications", *Embedded Hardware Design, Microprocessors and Microsystems*, Vol. 33, No. 3, May 2009, pp. 211-220.
- [11] M.Z. Hasan and S.G. Ziavras, "Resource Management for Dynamically-Challenged Reconfigurable Systems,"

12th IEEE Conf. Emerging Technology and Factory Automation, Patras, Greece, Sept. 25-28, 2007, 119-126.

- [12] Starbridge Systems, <http://www.starbridgesystems.com/hypercomputing/H-C-62/>.
- [13] The EDN Consortium, <http://www.eembc.org/>.
- [14] M.R. Guthaus, et al., "MiBench: A free, Commercially Representative Embedded Benchmark Suite", *4th IEEE Annual Workshop on Workload Characterization*, Austin, Texas, Dec. 2, 2001.
- [15] J. Noguera, and R.M. Badia, "Multitasking on Reconfigurable Architecture: Micro Architecture Support and Dynamic Scheduling", *ACM Trans. Embedded Computing Systems*, 3(2), 2004, 385-406.

Biography

MUHAMMAD ZAFRUL HASAN received the B.Sc. in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology. He received the Master of Electronic Engineering from Eindhoven University of Technology (The Netherlands) under a Philips postgraduate scholarship program. He subsequently held several faculty positions in an engineering college and in a university in Malaysia. He obtained the Ph.D. in Computer Engineering from New Jersey Institute of Technology. He was awarded the NJIT Hashimoto Fellowship in the academic year 2005-06. He is currently an Assistant Professor of Engineering Technology and Industrial Distribution at TAMU. His research interests include the design, implementation, and testing of dynamically reconfigurable computing systems, performance evaluation of computer architectures, and behavioral synthesis and testing of digital systems. He may be reached at hasan@entc.tamu.edu.

SOUND SOURCE LOCALIZATION EMPLOYING POLAR DIRECTIVITY PATTERNS OF BIDIRECTIONAL MICROPHONES

Vijay Varada, The University of Toledo; Hong Wang, The University of Toledo; Vijay Devabhaktuni, The University of Toledo

Abstract

This paper introduces a novel-sound-source localization technique using an array of bidirectional microphones. Specifically, the proposed technique exploits the polar directivity pattern of three bidirectional microphones to provide complete 360° localization coverage. The feasibility of this approach has been tested through both simulations and experiments. An algorithm was developed specifically for the proposed method. This technique reduces the total number of microphones required to provide similar coverage to the recently introduced unidirectional microphone-based-localization technique. At the same time, the overall spatial displacement requirements for the sensors were significantly reduced compared to the existing time-domain and phased-array methods.

Introduction

Sound-source localization (SSL) may be defined as identifying the location of a detected sound source. Sound-source localization techniques are usually passive with broad applications. For example, in warfare, SSL is used to enhance the patrol capability of troops [1], [2]. In conference meetings, SSL is used to direct and zoom cameras towards the speaker for a better viewing experience. In teleconferencing, it helps by allowing the network bandwidth to be more efficient [3], [4]. Sound-source localization may also be very important in humanoid robots and hearing aids. Finally, SSL can be employed in nocturnal activities such as night-hunting.

Due to its widespread applications, SSL has been an interesting topic for researchers over the past two decades, and various SSL techniques have appeared in literature [5-9]. In general, these techniques can be classified into three main categories, namely, beam-forming, time-delay, and spectral analysis [10-13]. In beam-forming, the microphone signal is steered in all possible directions to maximize the output [14], [15]. In time-delay methods, the time-of-arrival of the sound signal on different sensors is used for localization. In spectral analysis, spatial-spectral correlation among the received signals (usually narrowband) is used for source localization. Hybrid methods that combine beam-forming and time-delay have also been proposed [10], [16], [17]. These techniques, however, are limited by the very large spatial-

displacement requirements of the sensor, and, therefore, may not be practical in confined environments. Recently, a new approach exploiting the polar directivity pattern (PDP) of unidirectional microphones has been proposed [1]. This approach allows reduced sensor spacing compared to the prevalent methods. However, it provides limited SSL coverage in the plane of the microphone sensor array.

This paper introduces a new SSL method based on the use of bidirectional microphones. The proposed method provides complete 360° SSL coverage, while maintaining the reduced spatial-displacement requirements for the sensors. The unidirectional microphone-based SSL technique is discussed in Section 2, followed by the proposed approach in Section 3. Section 4 discusses the simulation results, followed by Sections 5 and 6, which present the experimental setup and results. Lastly, the conclusions are given in Section 7.

Prevalent SSL Technique Based on PDPs of Unidirectional Microphones

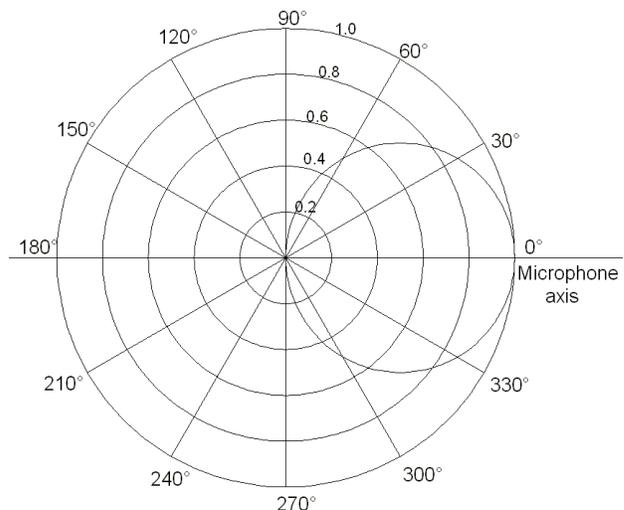


Figure 1. General polar directivity pattern of a first-order unidirectional microphone

The sensitivity/response of directional microphones has a strong dependence on the incidence angle of the sound signal described by equation (1) as

$$N = P + G * \cos \theta, \quad (1)$$

where N is the sensitivity of the microphone, P is the pressure on the diaphragm of the microphone caused by the incident sound wave, and θ is the angle of incidence of the acoustic signal. G is the gradient and is defined as the difference of pressure on either side of the diaphragm [18]. The general polar-directivity pattern of a first-order unidirectional microphone can be plotted as in Figure 1. As can be seen, the microphone gives the maximum output for $\theta = 0^\circ$, while for $90^\circ < \theta < 270^\circ$, the microphone output is zero.

The normalized output (N) of a directional microphone provides the angle of incidence of a detected signal in the PDP. Each normalized value can be geometrically represented as a circle of radius N_i (Figure 1). This circle intersects the PDP of the microphone giving rise to two angles for each normalized output. In the unidirectional microphone-based SSL technique [1], three microphones are arranged in a circular array. Two lines, corresponding to each normalized microphone output, are drawn from each microphone. Triangulation is then used to localize the sound source. However, this technique does not provide 360° coverage.

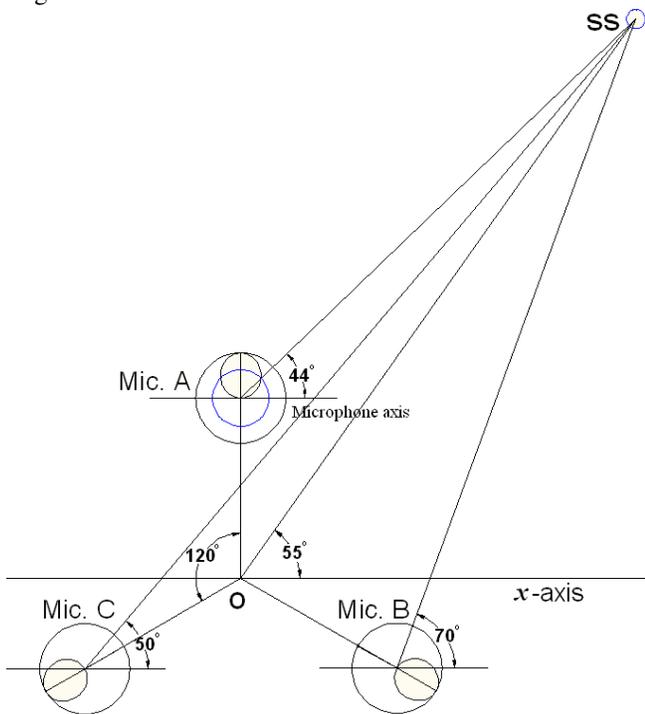


Figure 2. Localization using unidirectional microphones oriented at angles $90^\circ/330^\circ/210^\circ$

This disadvantage can be illustrated using the example shown in Figure 2. Here, three unidirectional microphones (labeled A, B, and C) are spaced equally on a circular path such that their axes are directed to 90° , 330° , and 210° with respect to the x -axis. A sound signal originating from a sound source (SS) located at an angle of 55° with respect to origin O will arrive at angles 44° , 70° and 50° to the micro-

phones A, B, and C, respectively. Using these values, one can find θ_A , θ_B and θ_C , which are the angles of the sound signal with respect to the microphone axes as shown in Figure 2. By using these angles, one can get the normalized values. As can be deduced from Figure 2, microphone A has some output, but microphones B and C produce zero output, i.e. they do not pick-up any signal. This means that localization cannot be performed for this particular source location, which can be referred to as a blind spot. To reduce the number of such blind spots, it is recommended to use more than three microphones in the array [1].

Proposed Approach

Bidirectional microphones have double ‘listening capacity’ and provide more information than unidirectional microphones [18]. This paper proposes using three bidirectional microphones, which have two lobes in their PDPs to remove the blind spots. Such an approach could make the system more robust. As will be shown, the proposed approach provides complete 360° localization with no blind spots. Substituting $P = 0$ and $G = 1$ for the bidirectional microphones in equation (1) [18], the polar equation becomes

$$N = \cos \theta. \quad (2)$$

The corresponding bidirectional PDP is plotted in Figure 3.

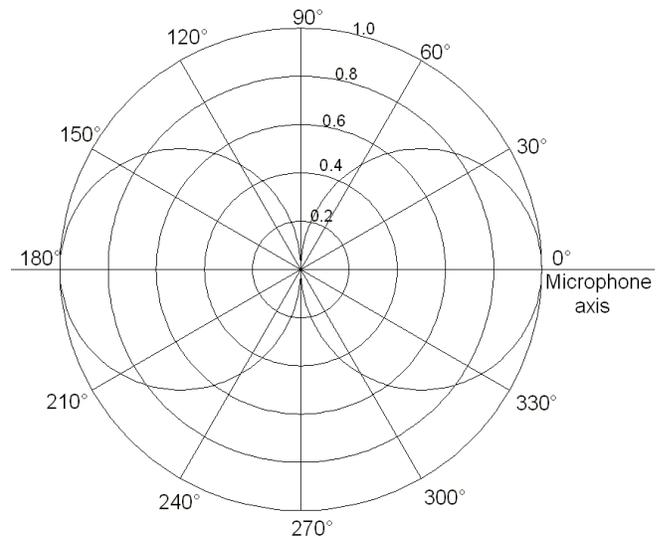


Figure 3. General PDP of a bidirectional microphone

For illustration of the proposed technique, consider Figure 4 showing three bidirectional microphones A, B, and C. Once again, the microphones are equally spaced and are located in a circular arrangement at 90° , 330° , and 210° , respectively.

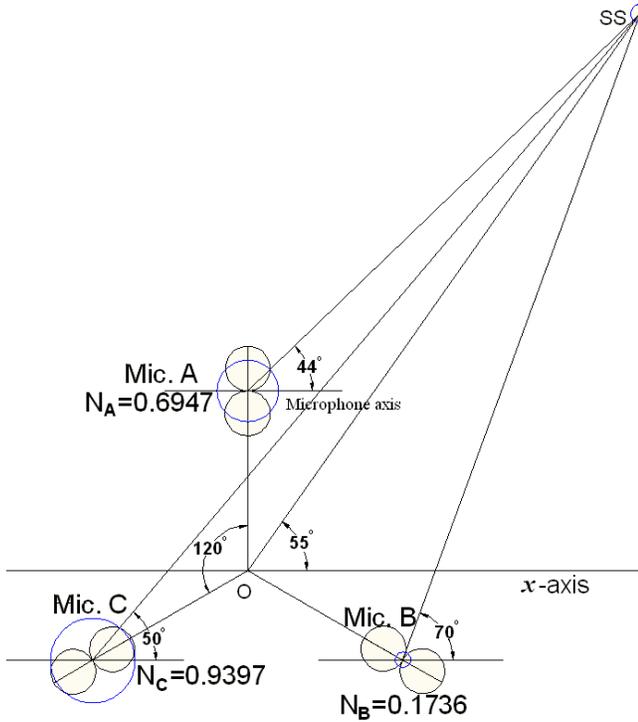


Figure 4. Localization using bidirectional microphones oriented circularly at angles 90°/330°/210°

Let normalized outputs from microphones A, B, and C be

$$N_A = \cos \theta_A, \quad (3)$$

$$N_B = \cos \theta_B, \quad (4)$$

$$N_C = \cos \theta_C, \quad (5)$$

where θ_A , θ_B and θ_C are angles with respect to the microphone axes. As can be inferred from Figure 3, corresponding to each normalized value, N_i ($i = A, B, \text{ or } C$), there are four angles in a bidirectional PDP. For each normalized output, four straight lines can be extended from the center of each microphone. Since the lobes are symmetrical, these lines connect at the microphone centers such that there are only two lines crossing the microphone center. The intersection of these lines leads to the localization of sound sources using the algorithm of Figure 5. The corresponding pseudo-code is presented in Table 1.

Simulation Results

For validation of the proposed approach, consider Figure 4, where three bidirectional microphones are arranged in a circular array. Each microphone is located at a distance of 4 units from the origin O , with their centers at $(0, 4)$, $(3.4641, -2.0)$, and $(-3.4641, -2.0)$, respectively. The sound source, SS,

is located at $(8.7650, 12.4053)$. By using equation (2), the normalized outputs (N_A , N_B , and N_C) from the microphones are 0.6947, 0.1736, and 0.9397. Corresponding to these normalized values, the potential incidence angles with respect to of the x -axis are, 44° and 136° at microphone A, 50° and 70° at microphone B, and 10° and 50° at microphone C.

The algorithm in Figure 5 was employed to trace the lines along the potential angles of incidence and to calculate the intersection points. Using all possible combinations, the triangle with the smallest area was determined. In practice, the instrument noise and acoustic propagation variations lead to triangles with finite areas.

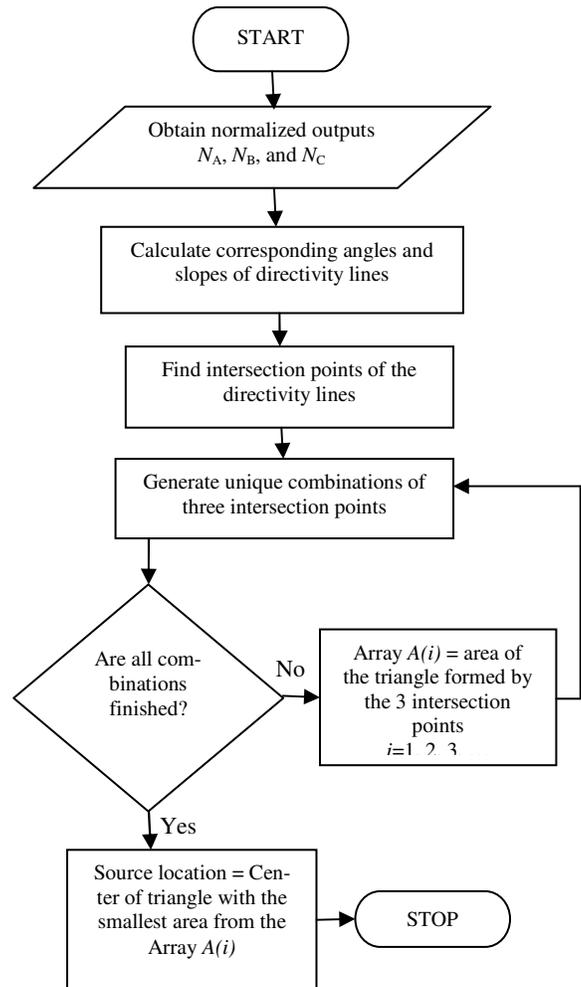


Figure 5. Flowchart of the proposed algorithm

However, since this simulation did not account for these imperfections, the lines intersected exactly at $(8.7650, 12.4053)$ which is, in fact, the location of the SS. A plot depicting the directivity lines (A_1, A_2 ; B_1, B_2 , and C_1, C_2) and their intersections is shown in Figure 6.

To check for the coverage of the proposed technique, the simulation was repeated with a 360° sweep of the SS placed at a distance of 12 units from the origin O . The results plotted in Figure 7 show complete 360° coverage with no null points.

Table 1: Pseudo-code of the proposed approach

<p>Step 1: Let the normalized outputs of the microphones be $N_A, N_B,$ and N_C. Find the angle of incidence θ with respect to the microphone axis using</p> $N_i = \cos \theta_i,$ <p>where $i \in [A, B, C]$.</p> <p>Step 2: Calculate the angles of the directivity lines $\alpha_A, \beta_A,$ $\alpha_B, \beta_B,$ and α_C, β_C using</p> $\alpha_i = \theta_i + \phi_i \text{ and } \beta_i = \theta_i - \phi_i,$ <p>where $i \in [A, B, C]$</p> <p>and ϕ_i is the orientation of the microphone.</p> <p>Step 3: Draw two lines from each microphone center:</p> <p>Line 1: $(y - y_i) = \tan \alpha_i (x - x_i),$</p> <p>Line 2: $(y - y_i) = \tan \beta_i (x - x_i),$</p> <p>where $i \in [A, B, C]$.</p> <p>Step 4: Solve pairs of these lines to obtain an array of intersection points P.</p> <p>Step 5: The center of the smallest triangle formed by any three points $\in P$ is the source location.</p>

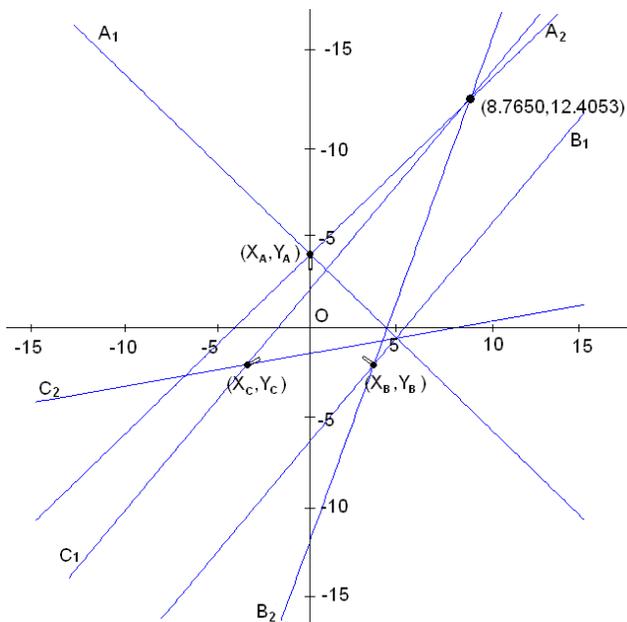


Figure 6. Localization using bidirectional microphones

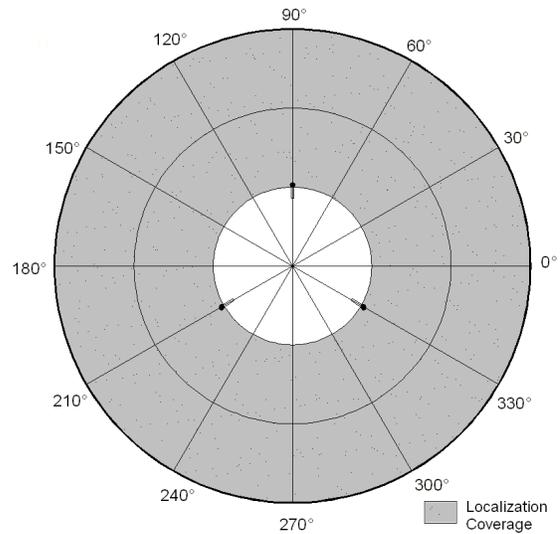


Figure 7(a). Localization sweep using bidirectional microphones

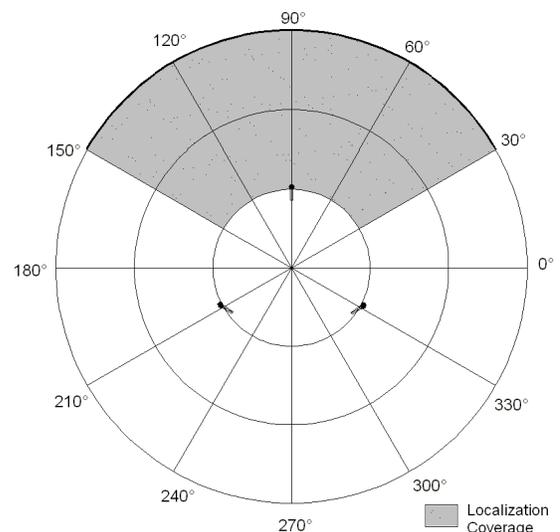


Figure 7(b). Localization sweep using unidirectional microphones

Similar simulations were repeated for the unidirectional microphones arranged at 90°/60°/120° (Figure 7), since the 90°/330°/210° arrangement proved unfeasible due to blind spots (Figure 2).

The plot in Figure 7 shows that the use of three unidirectional microphones provides only 120° coverage. It does not provide full 360° coverage with just three microphones. The proposed method provides complete coverage while keeping the number of microphones low. While the simulation results show the advantages of using the proposed method, experimental results discussed in the next section prove its feasibility.

Experimental Setup

A. Hardware and Software

The experimental setup consisted of three bidirectional microphones, model MXL R144. The manufacturer's frequency response is flat in the range 800-3000Hz. These microphones were connected to an oscilloscope by using XLR to BNC cables. A sound-generation software was developed using visual basic to produce sounds of desired frequencies. The readings taken from the oscilloscope were manually fed into a MATLAB program, which processes the data using the algorithm developed specifically for bidirectional microphones, and points to the location of the sound source with a graph diagram. Each experiment was repeated for 800Hz, 1000Hz and 1200Hz frequencies.

B. Polar Directivity Patterns

Measurements for 360° were taken by rotating the microphones at discrete steps of 10° to find the polar directivity patterns of the microphones. Each microphone was measured for frequencies of 800Hz, 1000Hz and 1200Hz) and, for each frequency, the distances of 6ft., 8ft. and 12ft. between the microphone and SS were considered. The measurements for all three microphones were found to be the same. Figures 8, 9 and 10 show the PDPs obtained from the measured data.

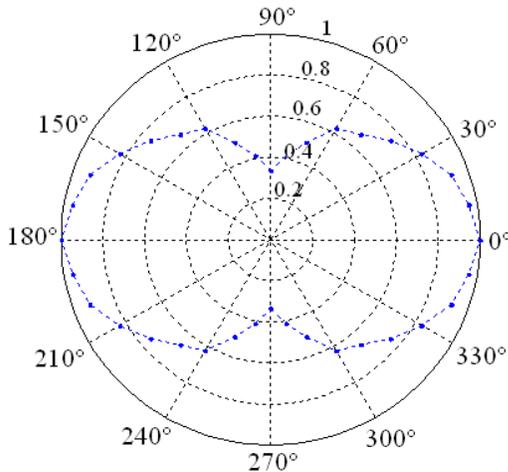


Figure 8(a). PDP at 800Hz and 6 feet from the source

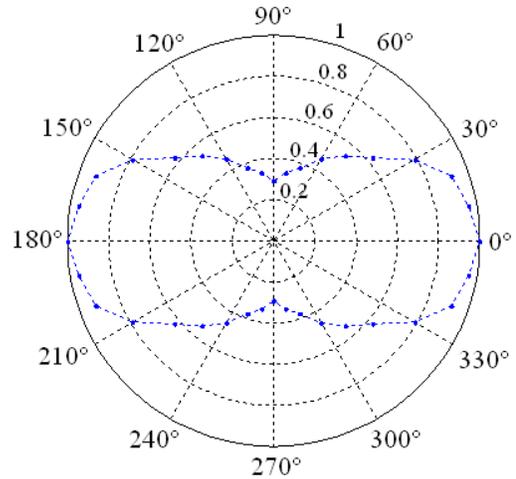


Figure 8(b). PDP at 800Hz and 8 feet from the source

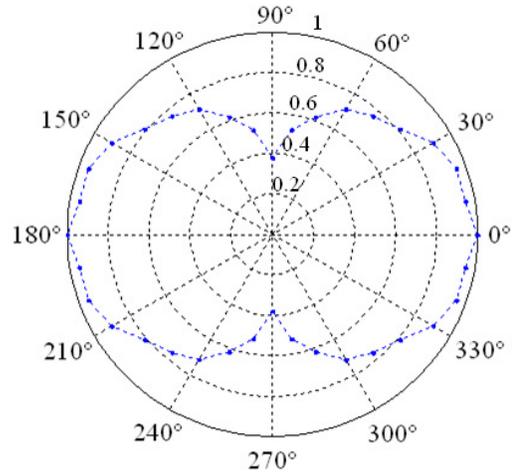


Figure 8(c). PDP at 800Hz and 12 feet from the source

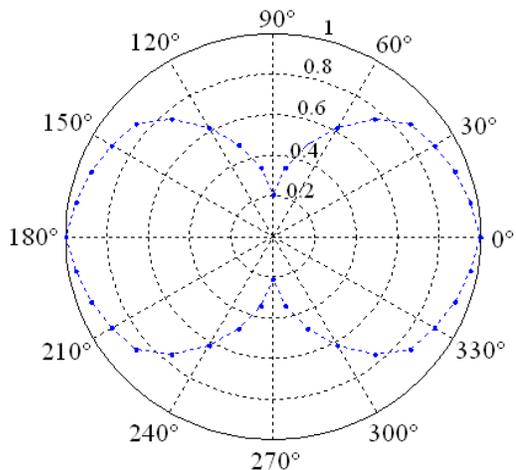


Figure 9(a). PDP at 1000Hz and 6 feet from the source

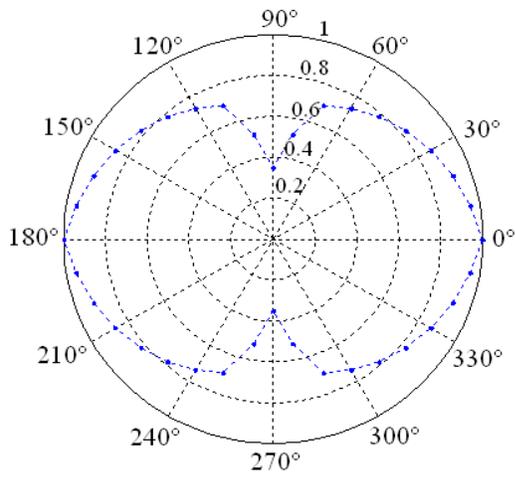


Figure 9(b). PDP at 1000Hz and 8 feet from the source

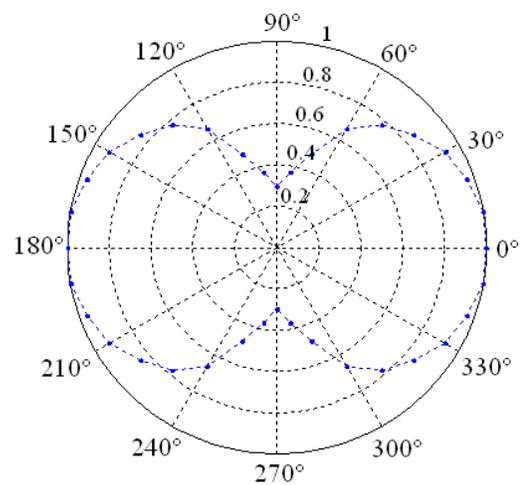


Figure 10(b). PDP at 1200Hz and 8 feet from the source

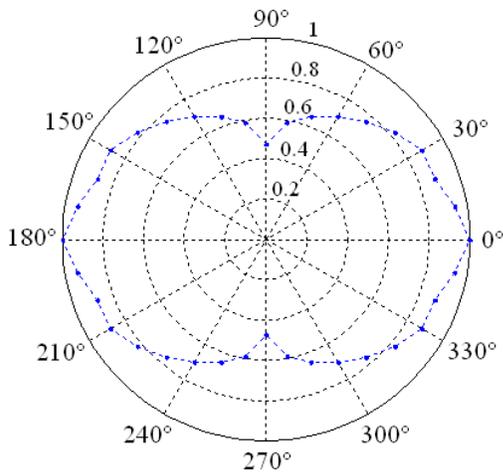


Figure 9(c). PDP at 1000Hz and 12 feet from the source

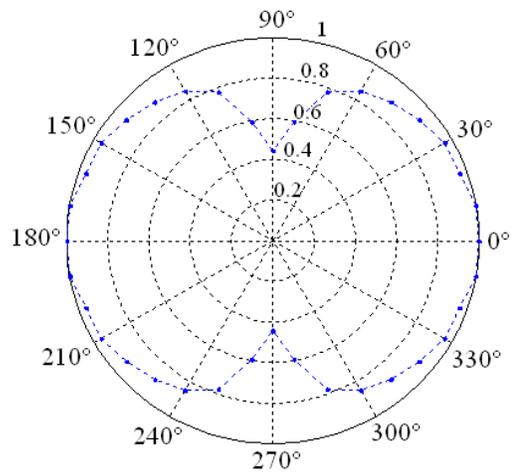


Figure 10(c). PDP at 1200Hz and 12 feet from the source

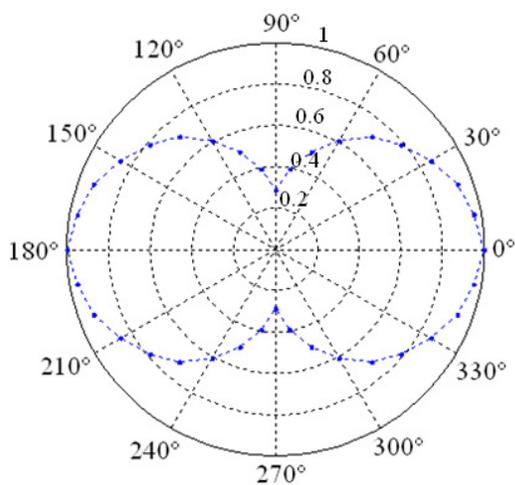


Figure 10(a). PDP at 1200Hz and 6 feet from the source

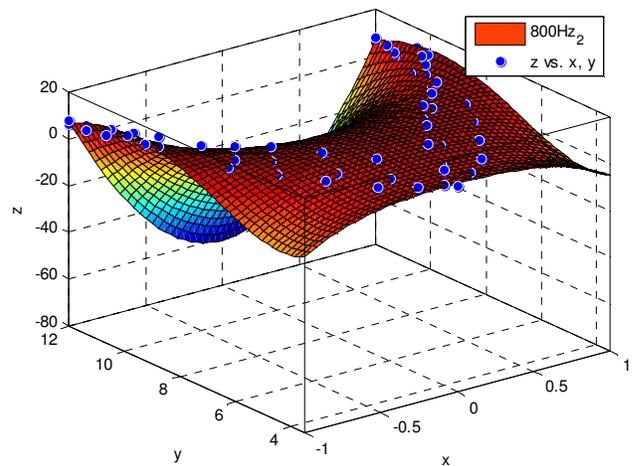


Figure 11. Surface fit using the measured values for 800Hz frequency

Using these results, the aim was to find a fitting equation. By finding the fitting equation, one can generalize it for any distance between the microphone and the SS. To demonstrate this, the surface-fitting toolbox in MATLAB was used. Figure 11 illustrates the surface plot generated for a frequency of 800Hz by taking the data that has been experimentally measured into consideration.

Similar plots for 1000Hz and 1200Hz were generated and the following equations (6), (7), and (8) were obtained for each of the three frequencies (800Hz, 1000Hz, and 1200Hz, respectively).

$$z = 80.18 - (46.55)y + (20.14)x^2 + (10.34)y^2 - (12.94)x^2y - (0.8883)y^3 - (6.474)x^4 + (1.454)x^2y^2 + (0.02239)y^4 \quad (6)$$

$$z = -17.26 + (14.95)y + (30.09)x^2 - (2.886)y^2 - (4.126)x^2y + (0.2208)y^3 - (3.434)x^4 + (0.1986)x^2y^2 - (0.006079)y^4 \quad (7)$$

$$z = -58.73 + (38.16)y + (37.61)x^2 - (8.161)y^2 - (22.98)x^2y + (0.7862)y^3 - (48.31)x^4 + (1.949)x^2y^2 - (0.02798)y^4 \quad (8)$$

With these equations, one can now perform experimental evaluations for various sound-source positions which will be discussed in the next section.

Experimental Results

Equation (1), which was used for simulation earlier, is now replaced with equations (6), (7) and (8) that were extracted from the experimental patterns to localize the SS.

The microphone arrangement was similar to the previous arrangement used for simulation. Here, a few more cases have been considered by decreasing the aperture of the microphone array to show that this method would be useful for making reduced-size sound localizers. First, the microphones were arranged in a circular fashion two feet from the center. Then the distance was reduced to 1 foot and then to 0.5 feet. Each of these arrangements was then tested for three different source locations. The entire process was repeated for the 800Hz, 1000Hz, and 1200Hz frequencies.

Figures 10, 11, and 12 show the experimental results. The three black dots A, B, and C represent the three bidirectional microphones and the red dot O is the origin. The red, green and blue pairs of lines are the directivity lines of the three microphones while the small black squares represent the experimental location of the sound source S_e .

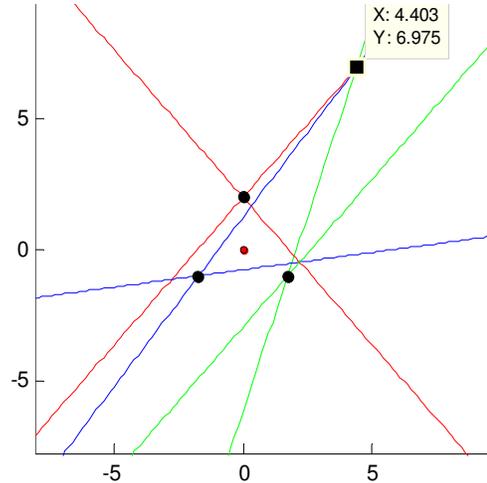


Figure 10(a). Experimental results for Q_{2f} for 800Hz

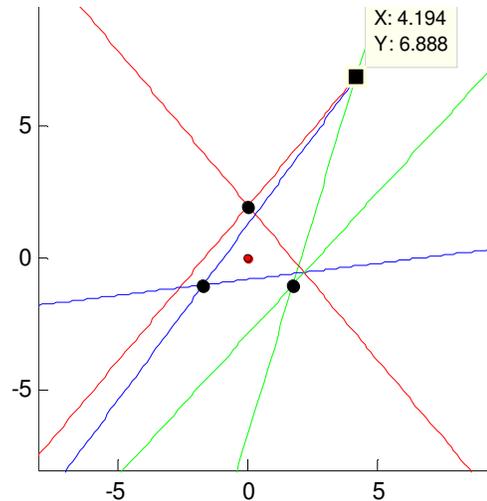


Figure 10(b). Experimental results for Q_{2f} for 1000Hz

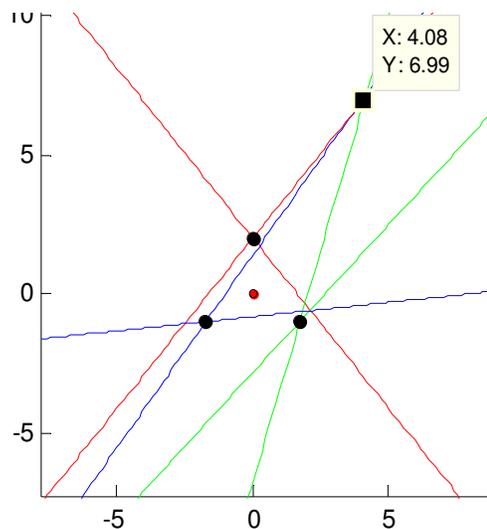


Figure 10(c). Experimental results for Q_{2f} for 1200Hz

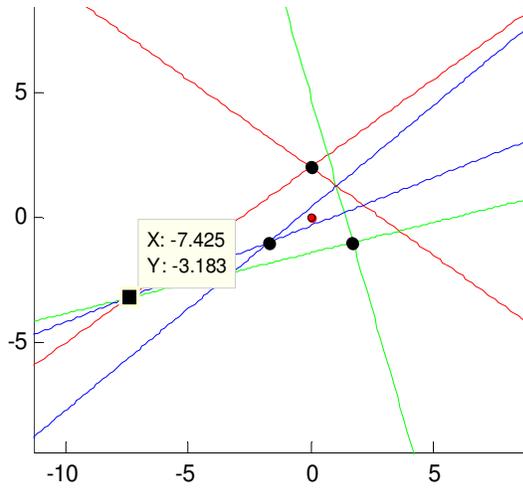


Figure 11(a). Experimental results for R_{2f} for 800Hz

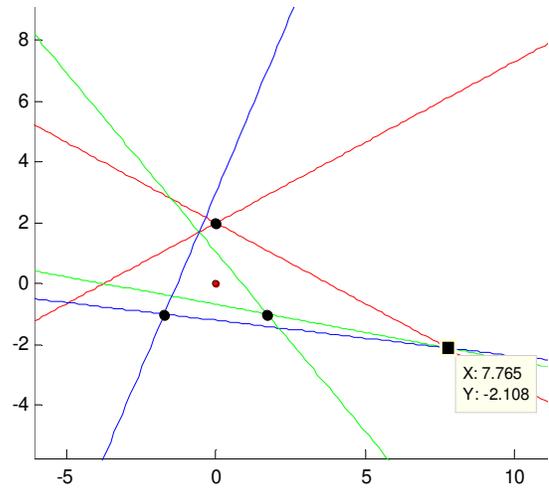


Figure 12(a). Experimental results for S_{2f} for 800Hz

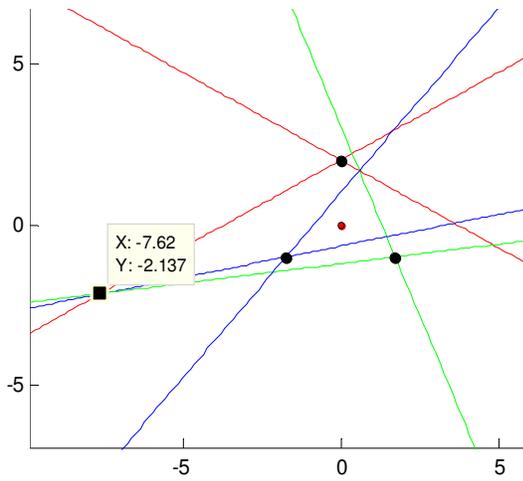


Figure 11(b). Experimental results for R_{2f} for 1000Hz

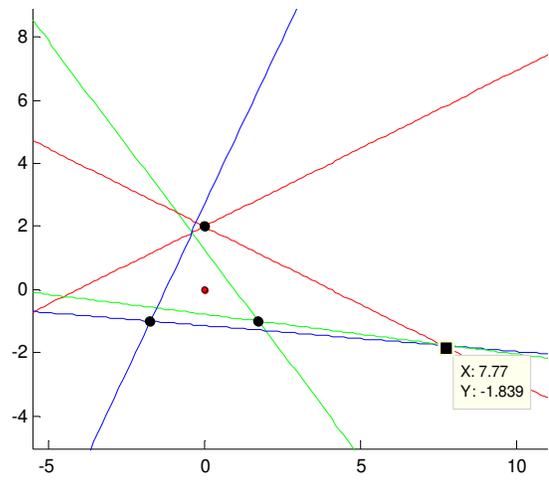


Figure 12(b). Experimental results for S_{2f} for 1000Hz

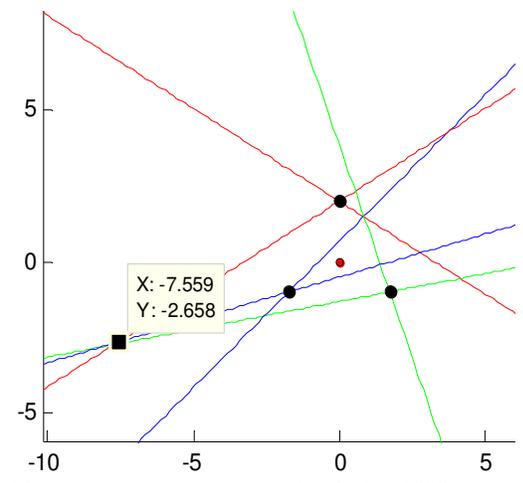


Figure 11(c). Experimental results for R_{2f} for 1200Hz

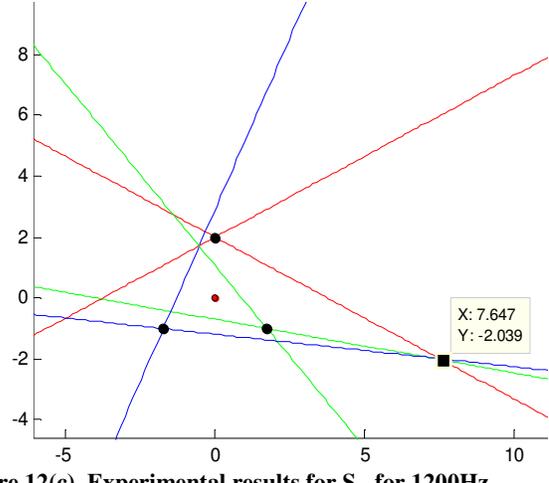


Figure 12(c). Experimental results for S_{2f} for 1200Hz

Table 2. Comparison of experimental results with actual source positions

Source Location	Experimental Location		
	800 Hz	1000 Hz	1200 Hz
$Q_{2f}(4.4105, 6.8589)$	(4.403, 6.975)	(4.194, 6.888)	(4.08, 6.99)
$R_{2f}(-7.5473, -2.6339)$	(-7.425, -3.183)	(-7.62, -2.137)	(-7.559, -2.658)
$S_{2f}(7.0396, 2.0503)$	(7.765, -2.108)	(7.77, -1.839)	(7.647, -2.039)

Table 2 compares the experimental results obtained for 800 Hz, 1000 Hz, and 1200 Hz frequencies with the actual sound-source location. The above results show that the sound-source positions obtained from the experimental values are very close to the actual positions. The small differences associated with the measured source positions could have resulted from human errors.

The same approach was used to test the reliability of the system by moving the microphones closer. This was tested at distances of 1-foot and 0.5-foot distances from the center to the microphone. The results were similar to the 2-foot arrangement. This shows that the technology is not sensitive to distance.

Certain sound-localization applications in real-life require the sensors to be placed far apart rather than clustering them into a single device. Since the proposed technique employs bidirectional microphones, which are able to “listen” from front and back sides, the “far apart” arrangement is possible, and would require modifications in the proposed algorithm. The current algorithm does not take into account the area between microphones, considering that the aperture is small and it is impractical to have a sound source at that position, i.e. inside the localizing device. If the microphones were to be placed far apart, the area would increase. Accordingly, the algorithm must be revised so that the area gets factored. From yet another perspective, since the aim of the paper is to pave way for miniaturized portable localizers, the aspect of large-aperture testing was not a part of our experiments.

Conclusions

This paper presents a novel sound-source-localization technique based on bidirectional microphones. In contrast to the technique relying on unidirectional microphones, the proposed technique provides complete 360° coverage with no blind spots in the plane of the microphone array. Experiments with different arrangements of microphones and different sound-source positions have been conducted. The sound-source positions have been located by intersecting all possible sound directions identified through the bidirectional polar patterns. The experimental results prove the feasibility of this approach. There was no difference in the results when

the microphones were moved closer to each other. It was seen that this technique works for low-aperture microphone arrays, making the system smaller and more portable. This shows that this kind of technology gives rise to a new generation of reduced-size sound localizers. The small differences associated with the measured source location could have been reduced if the microphone and sound source were moved/rotated with a more accurate mechanism instead of using a manual approach.

This study demonstrated the feasibility of the proposed approach which has 360° coverage using just three microphones as opposed to six microphones [1]. Also, this does not have any limitations on the microphone-array aperture like the time-delay approaches.

Future work will involve extending this technique to three-dimensional problems. The possibility of source localization using two microphones could also be evaluated, which could lead to even smaller localizing devices at a reduced cost. The plan is to build a number of hardware prototypes for demonstrating and fine tuning the proposed approach.

Acknowledgements

We gratefully acknowledge Dr. Samir Hefzy (Associate Dean of Graduate Studies and Research Administration, The University of Toledo) for providing financial assistance to Vijay Varada, which made this work possible.

References

- [1] P. Rizzo, G. Bordoni, A. Marzani, and J. Viperman, “Localization of sound sources by means of unidirectional microphones,” *Meas. Sci. Technology*, 20 055202, 2009.
- [2] D. K. Wilson, B. M. Sadler and T. Pham, “Simulation of detection and beamforming with acoustical ground sensors,” *Proc. SPIE, Seismic and Acoustic Unattended Ground Sensors I*, 2002, vol. 4743, pp. 50–61.
- [3] Y. Huang, J. Benesty, and G.W. Elko, “Passive acoustic source localization for video camera steering,” *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2000, vol. 2, pp. II909-12.
- [4] Y. Rui, D. Florencio, W. Lam, and J. Su, “Sound source localization for circular arrays of directional microphones,” *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2005, vol. 3, pp. 93-6.
- [5] M. S. Brandstein and H. F. Silverman, “A practical methodology for speech source localization with microphone arrays,” *Comput. Speech Lang*, 1997.

-
- [6] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," *J. Acoust. Soc. Am.*, 2000.
 - [7] S. Bourennane and A. Bendjama, "Locating wide band acoustic sources using higher order statistics," *Appl. Acoustics*, 2002.
 - [8] G. L. Reid and E. Milios, "Active stereo sound localization," *J. Acoust. Soc. Am.*, 2003.
 - [9] H. Liu and E. Milios, "Acoustic positioning using multiple microphone arrays," *J. Acoust. Soc. Am.*, 2005.
 - [10] J. Merimaa, "Application of a 3-D microphone array," *Proc. 112th Conv. Audio Eng. Soc.*, 2002, Paper 5501, pp. 1-10.
 - [11] X. Bian, G.D. Abowd, and J.M. Rehg, "Using sound source localization in a home environment," *Pervasive Comput.*, 3468, pp. 19-36, 2005.
 - [12] H. Atmoko, D.C. Tan, G.Y. Tian, and B. Fazenda, "Accurate sound source localization in a reverberant environment using multiple acoustic sensors," *Meas. Sci. Technology*, vol. 19, 024003, 2008.
 - [13] B. Kwon, Y. Park and Y-S. Park, "Sound source localization using the compensation method in robot platform," *Proc. IEEE Int. Conf. Control Autom. Syst.*, 2007, pp. 1911-4.
 - [14] X. Bian, J.M. Rehg, and G.D. Abowd, "Sound Source localization in domestic environment," *GVU Center, Georgia Inst. of Technology*, 2004.
 - [15] G. W. Elko, T. C. Chou, R. J. Lustberg and M. M. Goodwin, "A constant-directivity beamforming microphone array (A)," *J. Acoust. Soc. Am.*, 1994.
 - [16] H-D. Kim, K. Komatani, T. Ogata, and H.G. Okuno, "Design and evaluation of two-channel-based sound source localization over entire azimuth range for moving talkers," *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2008, pp. 2197-2203.
 - [17] A. Ikeda, H. Mizoguchi, Y. Sasaki, T. Enomoto, and S. Kagami, "2D sound source localization in azimuth and elevation from microphone array by using a directional pattern of element," *Proc. IEEE Sensors*, 2007, pp. 1213-6.
 - [18] R. Streincher and W. Doodley, "The bidirectional microphone: a forgotten patriarch," *J. Audio Eng. Soc.*, vol. 51, no. 3, 2003.

HONG WANG received his Ph.D. degree in Computer Science from Kent State University, Kent, OH, in 2006. Currently, he is an Assistant Professor of Computer Science and Engineering Technology at the University of Toledo. His teaching and research interests include Parallel Computing, Bioinformatics, and Embedded System Design. Dr. Wang may be reached at hong.wang2@utoledo.edu.

VIJAY DEVABHAKTUNI received the B.Eng. degree in Electrical and Electronics Engineering and the M.Sc. degree in Physics both from the Birla Institute of Technology and Science, Pilani, Rajasthan, India, in 1996, and the Ph.D. degree in Electronics from the Carleton University, Ottawa, Ontario, Canada, in 2003. Currently, he is an Associate Professor in the EECS Department at the University of Toledo. His research interests include computer aided design (CAD), image/signal processing, modeling, RF/microwave design, optimization and wireless sensing. Dr. Devabhaktuni may be reached at Vijay.Devabhaktuni@utoledo.edu.

Biographies

VIJAY VARADA received the B.Tech. degree in Computer Science from the Jawaharlal Nehru Technological University, Hyderabad, India, in 2008, the M.S. degree in Engineering from the University of Toledo, Toledo, OH, in 2010. He may be reached at vijaykvarada@gmail.com.

MODELING AND VALIDATION OF AUTONOMOUS RENDEZVOUS AND DOCKING OF AIR BEARING VEHICLES

Amir Mobasher, Alabama A&M University; Paul Shiue, Christian Brothers University; Hossein Jamshidi, Alabama A&M University; Zhengtao Deng, Alabama A&M University

Abstract

A new mathematical model representing the motion and control of an air-bearing vehicle (Chase Vehicle) was developed. Closed form solutions for the motion of the vehicle are presented. Subsequently, equations are utilized to arrive at a simple equation that describes the thruster firing time as a function of the distance traveled and the number of thrusters being fired. A set of experiments was performed in which the two thrusters at the rear of the chase vehicle were fired for durations of 2 and 3 seconds, and the resulting motions were compared with those obtained for theoretical predictions. Results obtained via this analytical technique compared remarkably well with those obtained with the simulations performed via Simulink including the experimental results obtained in the laboratory. The comparison between the simulations and analytical results yielded an error less than one percent for all the cases considered. Both theoretical and numerical predictions were within 1% for all cases studied.

Introduction

The research for space exploration has renewed interest in developing technologies to include programs in robotic and manned systems. Rendezvous and Docking missions between two spacecraft have been one of the primary topics of interest for NASA since 1960. Recent evolution of these topics provides examples of the Hubble Space Telescope, Space Shuttle, Russian MIR Dockings, and the International Space Station (ISS) construction missions. A common similarity linking the current NASA mission philosophy and the very first Gemini docking mission is that at least one of these spacecraft has always been piloted by astronauts and supported by a virtual army of ground personnel. The advance of Rendezvous and Docking technologies from manual to automated capabilities is the objective of the Autonomous Rendezvous and Docking (AR&D) project. The reduction in the recurring cost of routine docking missions is essential. For the missions requiring operations due to long communication delays, AR&D becomes predominant. The AR&D project objectives are establishing design criteria, test facilities, procedures, simulation techniques, etc. The simulation techniques influence standardization of the

AR&D systems. Establishment of the test facilities and procedures support the development and verification of future systems prior to flight. Demonstrations of most of the objectives were presented through flight experiments, 6 Degrees of Freedom (6-DoF) Hardware-In-The-Loop (HITL) simulations, and digital test.

The AR&D is primarily comprised of the Chase Vehicle (CV), Target Vehicle (TV), and 3-Point Docking Mechanism (TPDM). These are necessary in one or more phases of an automated docking mission. Some of the AR&D system mission scenarios include: (1) autonomous phasing and rendezvous with a target spacecraft after the CV arrival in-orbit, (2) automated approach and departure maneuvers, and (3) automated "soft dock" with the TV. The AR&D system is able to meet all of the requirements without ground intervention while providing real-time monitoring capability. Presently, NASA has several developing missions that will require AR&D capability.

The Chase Vehicle has an on-board computer that performs hardware command, telemetry, guidance, navigation and control, collision avoidance maneuvers (CAMs) and system monitoring functions. A critical element to the system is a Video Guidance System (VGS). This system provides real-time data to the on-board computer determining the relative position and altitude between the active sensor that is mounted on the chase vehicle and the passive target that is mounted on the target vehicle. The VGS uses laser diodes to illuminate retro-reflectors in the target. Then a solid-state camera detects the return from the target, and a frame grabber and digital signal processor convert the video information into the relative positions and altitudes [1-4]. A new generation of video-based sensors called Advanced Video Guidance System (AVGS) was developed with improved performance and additional capability for longer range operations. The new design combines the sensor head and electronics module into one unit [3]. The Target Vehicle is equipped with a set of cones that align with the TPDM latches. A set of passive reflectors serve as the Global Positioning System (GPS) to the CV having a stabilized altitude. This hardware was successfully integrated and tested at NASA-MSFC FRL (Flight Robotics Laboratory) to verify the operational characteristics of the VGS in the low-earth-

orbit environment. The 3-Point Docking Mechanism performs the actual physical latching of the two spacecrafts.

The AR&D system was designed and tested in a 1-G (earth's surface) environment. The operation of the mission is accomplished by stepping through a set of automated maneuvers through which the CV moves towards the TV, leading to an eventual dock [5], [6]. The initial design and integration of the AR&D was completed in 1998 and tested in 1999 at NASA's Marshall Space Flight Center (MSFC). The facility is comprised of a 30 x 10m² flat floor which houses an air-sled with 3 degrees of freedom, a 3-DoF Chase Vehicle with associated hardware and software, and a 3-point docking mechanism. The air-sled weighs around 2500 pounds and is equipped with 10 mini thrusters, each capable of generating 4 pounds of thrust and can steer the vehicle in the fore, aft, and yaw orientations. Its purpose was to demonstrate rendezvous and docking of spacecraft in Earth orbit. Some of the latest innovations in Global Positioning System space navigation such as laser-sensor technologies and automated mission-sequencing algorithms were integrated into this system. Performance of the ground-based AR&D system was exceptional during the tests performed over a period of 6 months [6].

In this study, a new mathematical model for the motion of the 3-DoF ground-based Chase Vehicle in the Flight Robotics and Contact Dynamics Simulation laboratory at Marshall Space Flight center was developed. This mathematical model describes the motion of the vehicle as a function of time for the fore and aft positions. The model was then used to make a parametric study to correlate the thruster firing time with distance and the number of thrusters fired. As a result of this study, a two-step docking process was proposed. First, an assessment of the orientation of the CV relative to the target was achieved. Then, a prediction on the thruster firing time requirements was made and the chaser vehicle was brought within the vicinity of the three-point docking system. In the second step, on-board closed loop control algorithms were activated to perform the final docking. A set of experiments was performed in which the two thrusters at the rear of the air sled were fired for durations of 2 and 3 seconds after which the resulting motions were compared with those obtained for theoretical predictions as well as those obtained via simulations.

Methodology and Approach

The air-bearing vehicle (Chase Vehicle) used in the flight Robotics Lab at NASA-MSFC for the simulation of AR&D scenarios consisted of 3 air bearings responsible for lifting the vehicle, as well as ten thrusters, three-point docking cones and a laser range finder with associated software and hardware. Six compressed air tanks, embedded within the

vehicle, supplied the energy needed for lifting the vehicle as well as generating thrust in the fore and aft directions. Figure 1 depicts a two-dimensional view of the ground-based chaser vehicle with its associated components.

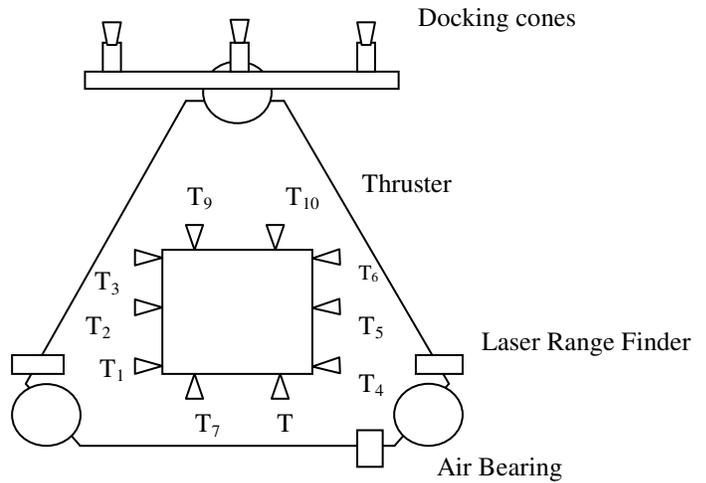


Figure 1. Schematic of the air-bearing vehicle (Chase Vehicle) used in the Laboratory setup. The vehicle consists of 3 air bearings responsible for lifting the vehicle which are supplied by three pressurized tanks, ten thrusters, three-point docking cones and laser range finder with associated software and hardware

The simplified model of the air-bearing vehicle is shown in the form of a free body diagram displayed in Figure 2. It has mass M , with the coefficient of friction μ , and is subjected to the thrust force of \mathfrak{S} . Writing the equation of motion for this system yields:

$$M\ddot{x} = \mathfrak{S} - \mu Mg \quad (1)$$

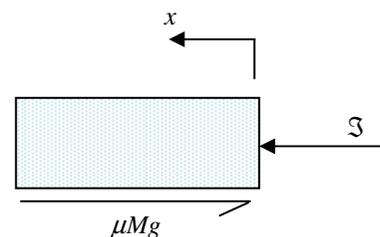


Figure 2. Free body Diagram of the air bearing vehicle with the forces acting on it

Due to the loss of mass (air) in the system, the mass of the air-bearing vehicle at a given time is given by

$$M = M_0 - \dot{m}t \quad , \quad (2)$$

where M_0 is the initial mass of the air-bearing vehicle inclusive of fuel, and $\dot{m}t$ is the mass of the fuel lost at a given time, t . The acceleration of the vehicle at a given time t is obtained as

$$\ddot{x} = \frac{\mathfrak{S}}{M_0 - \dot{m}t} - \mu g \quad (3)$$

Integrating equation (3) using initial condition $\dot{x}(0) = u_0$ along with simplification, velocity is obtained

$$\dot{x} = u_0 + \frac{\mathfrak{S}}{\dot{m}} \ln\left(\frac{M_0}{M_0 - \dot{m}t}\right) - \mu g t \quad (4)$$

Further direct integration of equation (4) applying initial condition $x(0) = x_0$ yields a non-linear expression for the position as:

$$x = \left[x_0 - \left(\frac{\mathfrak{S}}{\dot{m}}\right) \left(\frac{M_0}{\dot{m}}\right) \ln\left(\frac{M_0}{M_0 - \dot{m}t}\right) \right] + \left[u_0 + \frac{\mathfrak{S}}{\dot{m}} \ln\left(\frac{M_0}{M_0 - \dot{m}t}\right) + \frac{\mathfrak{S}}{\dot{m}} \right] \cdot t - \frac{1}{2} \mu g t^2 \quad (5)$$

Equations (3), (4), and (5) are valid as long as the thrusters are on. Once the thrusters are shut down (i.e., no external load is applied to the vehicle and $\dot{m} = 0$), the foregoing equations reduce to the following:

$$\text{acceleration: } \ddot{x} = -\mu g \quad (6)$$

$$\text{velocity: } \dot{x} = u_0 - \mu g t \quad (7)$$

$$\text{position: } x = x_0 + u_0 t - \frac{1}{2} \mu g t^2 \quad (8)$$

Equations (3) through (5) are solved for the case while thrusters are on and equations (6) through (8) are used when the thrusters are turned off. The CV goes to rest via friction. It must be noted that equations (6) through (8) are implicit within equations (1) through (4). A reasonable graphic representation of the position of the CV as a function of time is shown in Figure 3. A typical laboratory mission consists of a thruster firing time in the range 1 to 12 seconds during which the CV is accelerated. After that, all thrusters are shut down and the momentum carries the air-bearing vehicle forward. However, very low friction in the laboratory environment is taken into consideration for the deceleration period of the vehicle.

Equation (5) provides a relationship for the total distance traveled, d , as a function of thruster firing time, t , and the number of thrusters fired, n ; that is, $d = d(t, n)$. However, in a practical sense, the interest lies in finding an expression that describes a relationship between the firing time of the thrusters as a function of the total distance that the CV travels and the number of thrusters fired, that is $t = t(d, n)$.

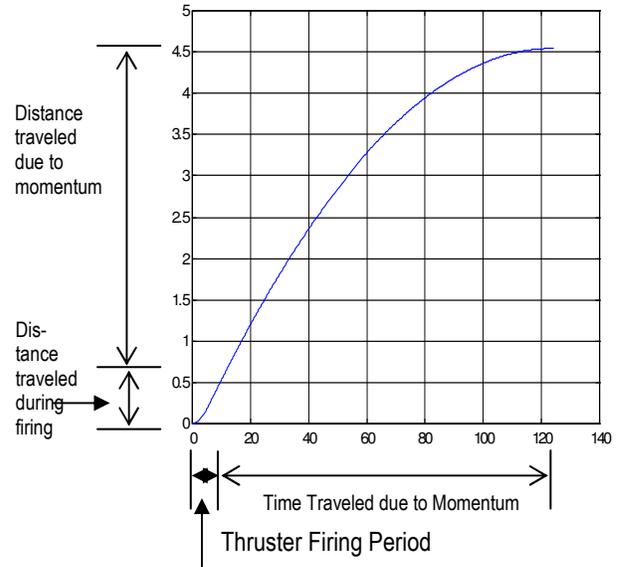


Figure 3. Typical motion of an air-bearing vehicle once the thrusters are fired for a certain period of time. The first-stage vehicle accelerates as long as thrusters are active. Once thrusters are shut down, the vehicle travels a certain distance and finally comes to rest due to friction

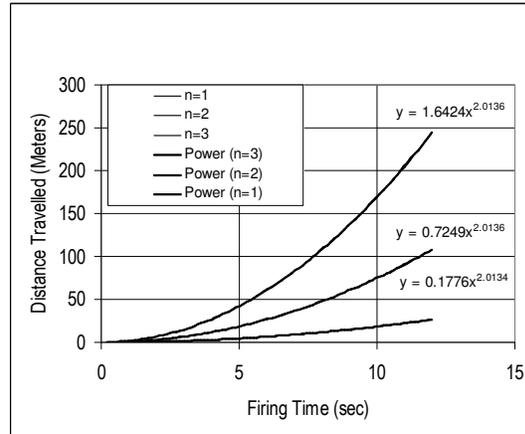


Figure 4. Family of curves for distance traveled as a function of the number of thrusters fired

To accomplish $t = t(d, n)$ a test matrix varying thruster firing time and number of thrusters (n) was established. In the test matrix, the thruster firing time ranges between 0.2 to 12 seconds and the number of thrusters varies from 1 to 3. Simulations were performed for each case, and the distance corresponding to the thruster firing time and number of thrusters was recorded. To establish the relationship for $d = d(t)$, a least square curve was passed through the resulting data points (distances). To find a correlation for the number of thrusters, a least square curve was again passed through

the coefficients of the individual set of equations, which are shown in Figure 4. The resulting equation for the individual set of equations was represented by $d = d(n,t)$. The form of this resulting equation was obtained as

$$d = 0.183n^2t^2 \quad (9)$$

which, upon simplification, yields

$$t = 2.337 \frac{\sqrt{d}}{n} \quad (10)$$

Equation (10) emphasizes the time required for firing n thrusters to carry the CV a total distance d either in the x or y direction. In practice, this equation may be employed to position the CV within the vicinity of the target area and then use the control algorithms for docking the CV.

Results and Discussions

In order to compare the validity of equation (10) Simulink was used to simulate the physical model and then employed to run a set of parametric tests for a varying number of thrusters from $n = 1, 2,$ and 3 for firing duration $t = 1$ through 8 seconds. The outcome of the test cases for the analytical results versus the simulations is presented in Table 1.

Table 1. Comparison of the distance traveled obtained via analytical and simulation results performed by Simulink for the thruster firing times of 1 through 8 seconds and number of thrusters varying from 1 to 3

Firing Duration (sec)	Number of thrusters Fired		
	$n = 1$	$n = 2$	$n = 3$
1	0.18 m (analytical)	0.73 m	1.64 m
	0.18 m (numerical)	0.74 m	1.60 m
2	0.73 m	2.93 m	6.59 m
	0.71 m	2.92 m	6.58 m
3	1.64 m	6.59 m	14.83 m
	1.60 m	6.58 m	15.01 m
8	11.72 m	46.71 m	-
	11.53 m	46.63 m	-

In all cases, the analytical results obtained via equation (10) compared remarkably well, within 1% error limit, with those obtained via the Simulink simulations. Figure 5 compares the results of the theoretical equations for displace-

ment given by equation (5) with those of the experimental results for the fore and aft displacements.

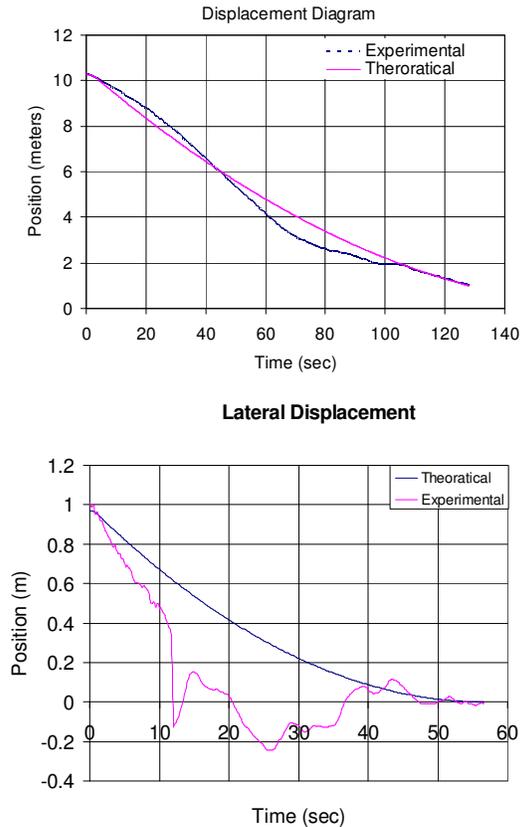
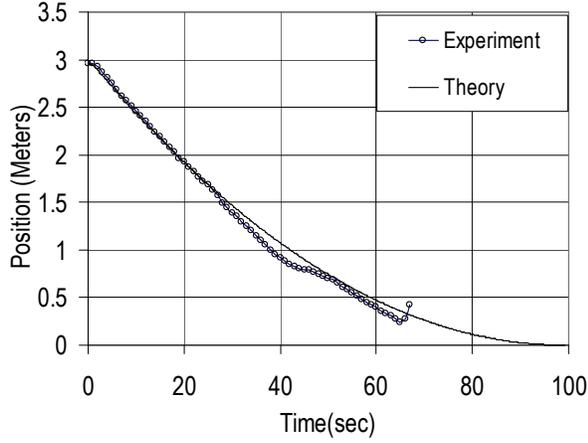


Figure 5. Comparison of theoretical and experimental results for (a) forward, and (b) lateral displacements as a function of time of the CV

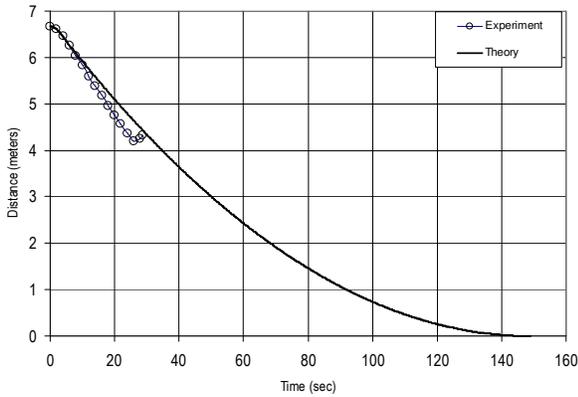
The control algorithm was utilized for zeroing in the air-bearing vehicle to the one-meter range. The experimental results for the fore displacement compared relatively well with the theoretical prediction. However, for the lateral (aft) position, experimental results deviated significantly from the theoretical prediction. To validate equations (5) and (10), a set of experiments was performed in which two thrusters at the rear of the air sled were fired for durations of 2 and 3 seconds. The resulting motion was compared with those obtained for the theoretical model.

Further examination of the theory and experiments was conducted for firing durations of 2 and 3 seconds. The plots are displayed in Figures 6(a) and 6(b), respectively. For the case of a 2-second firing time, the maximum distance traveled was predicted as 3m. The experiment and theory matched fairly well. It appears that for both cases, where thruster firing period increases, the theory and experiment compared somewhat better than those of smaller thruster firing periods. Each experiment was repeated three times;

however, due to the imperfections in the floor and other noises within



(a)



(b)

Figure 6. Comparison of path prediction between experimental and theoretical results for 2 seconds firing duration (a) and 3 second firing time (b)

the laboratory environment such as air movement caused from air conditioning ducts, it was difficult to draw meaningful average data from the three tests. Experimental data in this study proved to be highly sensitive to the environment. Other sources of error may be due to the friction coefficient used for the simulation and time delay used in the software program in firing the thrusters.

Further, to examine the experimental results (Figure 6-a) obtained in the laboratory with the analytical results obtained in theory, given by equation (5), the hypothesis statistical test was conducted.

In the case under investigation, the experimental results pertain to measurements of distance traveled by the sled for a firing duration of 2 seconds and analytical results implies the distance traveled by the sled for the aforementioned firing duration and calculated by equation (5). The hypothesis testing enables one to make a comparison between the performances of the means of experimental with analytical results and determine if the differences in the means are statistically significant. The paired t-test is the appropriate statistical test since the two populations, experimental and analytical, have the same related characteristic of interest which is time. Therefore, the following hypothesis was conducted:

- H_0 : There is no difference in the means of two related populations of experimental and analytical data
- H_1 : There is a difference

The above hypothesis can be represented as:

- H_0 : $\mu_D = 0$ (where $\mu_D = \mu_{\text{experimental}} - \mu_{\text{analytical}}$)
- H_1 : $\mu_D \neq 0$

where μ_D is the mean difference between the experimental and analytical results. The pair t-test was calculated as:

$$t = \frac{\bar{D} - \mu_D}{\frac{S_D}{\sqrt{n}}} \quad (11)$$

where \bar{D} is the average difference between the experimental and analytical results given by:

$$\bar{D} = \frac{\sum_{i=1}^n D_i}{n}$$

and

$$S_D = \sqrt{\frac{\sum_{i=1}^{94} (D_i - \bar{D})^2}{n-1}}$$

where S_D is the standard deviation of the differences. For a two-tailed test of the hypothesis with a given level of significance, α , you reject the null hypothesis if the computed t-value obtained in equation (11) is greater than the upper critical value (t_{n-1}) or if it is less than the lower critical value ($-t_{n-1}$). Based on the sample of 94 observations, taken within the same time frame for both experimental and analytical results, using the standard SPSS statistical package, Table 2 was obtained.

Table 2. Results of Pair t-test for Experimental Data Versus Analytical Data Using SPSS Statistical Package

Paired Samples Statistics

		Mean	N	Std. Deviation
Pair 1	EXPERIMENTAL	2.028060	94	0.486873006
	THEORETICAL	2.025049	94	0.476461643

Paired Samples Correlations

		N	Correlation	Sig
Pair 1	EXPERIMENTAL & THEORETICAL	94	1.000	.000

		Pair 1	
		EXPERIMENTAL & THEORETICAL	
Paired Difference	Mean	0.00301	
	St. Dev	0.0131	
	St Err Mean	N/A	
	95 % Confidence Interval of the difference	Lower	0.00033
Upper		0.0057	
t		2.228	
df		93	
Sig(2-tailed)		0.028	

The results in Table 2 reveal that the calculated t-value from equation (11) was $t = 2.228$. Based on the α value (level of significance), the following analysis was done:

α	Upper critical value	Lower critical value
0.05	1.9853	-1.9853
0.01	2.3662	-2.3662
0.005	2.6286	-2.6286

The null hypothesis was accepted (there is no difference between the mean performance of the experimental versus analytical data) as long as the t-value is between the upper and lower critical values. For $\alpha = 0.05$ the null hypothesis was rejected and it was concluded that the mean performance of the experimental results was significantly different than the theoretical results. However, for $\alpha = 0.01$ or smaller the null hypothesis was accepted since $-t_{n-1} = -2.3662 < t = 2.228 < t_{n-1} = 2.3662$ and concluded that the mean perfor-

mances of the experimental results is the same as the theoretical results. In addition, the RMS-error which measures the performance of experimental data was obtained as follow:

$$RMS = \sqrt{\frac{\sum (X_e - X_t)^2}{\sum (X_t)^2}} = \sqrt{\frac{0.566}{415.324}} = 0.0369$$

The result of RMS revealed the percentage of error or deviation of the experimental results from the theoretical results to be as low as $RMS = .0369$.

Conclusions

Analytical solutions obtained describe the position of the CV as a function of time and number of thrusters fired. A working equation was developed that defined the firing time requirement for the distance traveled. This equation compared well with the simulation results obtained via Simulink. The experimental data obtained for the 2- and 3-second firing times and 2 thrusters compared relatively well with the numerical predictions. The statistical analysis performed using SPSS software reinforced the validity of the theoretical equation versus the experimental at the α -less-than-0.01 level of significance.

Acknowledgements

The authors would like to acknowledge the contribution of the entire AR&D group at NASA Marshall Space Flight Center for this study. Thanks are due to Dr. Mike Freeman and Dr. Gerald Karr for their coordination and support of the Faculty Summer Internship Program leading to this study. Also thanks are due to Dr. M. A. Alim for useful discussion.

References

- [1] Howard, R. T., M. L. Book and T. C. Bryan, "Video-based sensor for tracking 3-dimensional targets," Proceedings of Spice Laser Radar technology and Applications, 4167, 242-251, 2000.
- [2] Howard, R. T., T. C. Bryan, M. L. Book and J. Jackson, "Active sensor system for automatic rendezvous and docking," Proceedings of Spice Laser Radar technology and Applications II, 106-115, 1997
- [3] Roe, F. D. and R. T. Howard, "The successful development of an automated rendezvous and capture (AR&C) system for the National Aeronautics and Space Administration," Proceedings of the 26th Annual American Astronautical Society Guidance and Control Conference, 2003
- [4] Tanner, E., S. Granade and C. A. Whitehead, "Autonomous rendezvous and docking sensor suite," Pro-

-
- ceedings of Spice Laser Radar Technology and Applications VIII, 5086, 329-339, 2003
- [5] Cruzen, C. A. and J. J. Lomas, "Design of the automated rendezvous and capture docking system," Proceedings AIAA International Space Station Service Vehicles Conference, 2000.
- [6] Cruzen, C. A., J. J. Lomas and R. W. Dabney, "Test results for the automated rendezvous and capture system," Proceedings of 23rd annual American Astronautical Society Guidance and Control Conference, 2000

Biographies

AMIR MOBASHER is an Associate Professor in the department of Mechanical Engineering at Alabama A&M University. He received his PhD from University of Alabama-Huntsville. Currently, he is a NASA Administrator's Fellow as part of the United Negro College Funds Special Programs (UNCFSP) working at NASA Marshall Space Flight center. His research spans the areas of Computational Fluid Dynamics, Biodynamics, Manufacturing Systems, Instrumentation and Control Systems and Autonomous Systems for Lunar Rovers. He may be reached at amir.mobasher@aamu.edu.

PAUL SHIUE is a Professor and Chair of the Mechanical Engineering Department at CBU. He received his B.S. from Tatung University in Taiwan and his M.S. and Ph.D. degrees from the University of Memphis. He is an associate member of the American Society of Mechanical Engineers and a professional member of the American Society for Engineering Education. Dr. Shiue is also a member of editorial advisory board of the International Journal of Engineering Education (IJEE), and he served as guest editor of a special issue in manufacturing engineering education (V20, N4, 5). He was five times a NASA/ASEE Summer Faculty Fellow at Marshall Space Flight Center. Currently, he is focusing on concurrent engineering and design through manufacturing and product realization processes. Dr. Shiue specializes in mechanical systems design and manufacturing, kinematics and dynamics, vibrations, and material testing. He was recognized as the Professor of the Year by Tau Beta Pi, Delta of Tennessee in the 1997-98 academic year. He also received the 1997 Featured Engineer for Excellence in Educating Engineers, Memphis Joint Engineering Council Achievement Award by CBU. He has been recognized in Who's Who Among America's Teachers in 1996, 2002, and 2005. He may be reached at pshiue@cbu.edu.

HOSSEIN JAMSHIDI is a full professor at Alabama A&M University, Normal, AL. He received his Ph.D. from The University of Alabama in Huntsville in Industrial and Systems Engineering. His areas of specialty are production

operations, mathematical programming, statistical analysis, and analysis of manufacturing systems. He has published many papers in these areas. He has conducted workshops for major companies such as Rolls-Royce/The Purdy Corp and Boeing/DACA program, Sponsored by U.S. Air Force. He has won many awards at the university, state, and national level. Dr. Jamshidi may be reached at hossein.jamshidi@aamu.edu

ZHENGTAO DENG is a professor in the department of Mechanical Engineering at Alabama A&M University. His research is focused on Advanced Space Propulsion and High Performance Computing. Dr. Deng may be reached at zhengtao.deng@aamu.edu.

CASE STUDY OF AN ADAPTIVE AUTOMATED HEALTH INSURANCE FRAUD AUDITOR

Fletcher Lu, University of Ontario Institute of Technology

Abstract

One of the significant potentials of encoding health records into an electronic format is as a vast resource that may be mined to find hidden relationships such as the task performed by auditors in their search for fraud. However, the simplicity of the idea belies the difficulty of this task. The key software engineering challenge involves extracting information from different sources despite being presented in different formats. An additional challenge is to merge this information with multiple fraud detection methods to take advantage of all the current detection techniques available. In this paper, a case study of a prototype software implementation of an automated fraud auditor is presented which was designed to replicate the investigative operation of human fraud auditors. The focus was on the adaptive design of the system. The implementation of this design on a set of real health insurance and hospital records, as well as a performance test on real audited data, demonstrated its improved efficiency over human auditor fraud case building.

Introduction

In 1999, Stead and Lorenzi [1] illustrated the need to link investment in Health Informatics to derived value from that investment in terms of improved public health, improved quality as perceived by consumers, and lowered costs. The automated fraud auditor presented here addressed the issue of lowering the costs of providing health care by reducing instances of health insurance fraud. A cost that Simborg [2] estimated to be between 3% and 10% of the total health care costs, which is a huge range illustrating the great amount of uncertainty in the amount of fraud that may go undetected.

Building a software application that will replicate the operation of a human fraud auditor, however, presented many challenges. The key challenge involved the core definition of fraud. The National Health Care Anti-Fraud Association (NHCAA) defines it as an intentional deception or misrepresentation made by a person or an entity with the knowledge that the deception could result in unauthorized benefits to that person or entity [3]. Thus, fraud requires two components:

1. deception, and
2. an unjustified gain or loss.

What these two components imply is that whatever fraud is perpetrated, it is generally hidden to some degree and some party must obtain an unjustified benefit or loss. Automated fraud-detection systems have been developed that use pre-defined sets of rules built by domain experts to uncover specific types of fraud. These pre-defined requirements limit their broad-based usability. More flexible methods capable of uncovering new patterns of fraud involve statistical machine-learning techniques that may be classified as either supervised or unsupervised. Supervised methods train on data that has been labeled by domain experts for fraud and then searches on general data for information that satisfies the fraud patterns learned from the training data. Unsupervised methods, in contrast, search for outliers/anomalies in the data. They potentially have a greater chance of uncovering new types of fraud as they are not restricted in any way to preset rules or trained data.

The software application developed in this study differs markedly from these previous fraud-detection tools in the fact that it may use these fraud detection tools to uncover potential cases of fraud. But it goes one step further by building a case of fraud around those potential cases, just as an auditor might do, by searching for related records to ascribe the unjustified gain or loss to some party. This new system, then, may best be described as a fraud-detection case builder, unlike any commercial system currently available in its case-building objective.

Just as human auditors may use any and all possible fraud/anomaly detectors, the automated system developed here may do so as well. The utility of this mechanism is in

- automating the task of human auditors who have to run multiple detectors and then manually explore the results to build the fraud case;
- speeding up the searching of the possibly vast amounts of data related to an outlier that must be sifted through in order to link records together that build a case for fraud; and
- relating data across different database tables with different data representation formats for the same information.

These challenges may be compounded by the large number of potential fraud cases that may need to be investigated depending on the magnitude of the original search space that the outliers were drawn from. In addition, healthcare data-

bases are continuously growing with new data continuously presenting new potential cases of fraud.

The system developed in this study uses a form of semi-supervised learning known as reinforcement learning [4] to allow it to have the flexibility to take information from multiple detectors and search data from multiple databases that could be in numerous data formats. The authors provide a background look at reinforcement learning, the various current fraud-detection techniques, a detailed explanation of the system’s design, and a case study of the proposed system’s operation on data from a real insurance provider.

It should be emphasized that this system is a fraud case builder paralleling the tasks of human fraud auditors. This means that the system can take information from other fraud detection programs to build its case for fraud. Thus, it should operate at least on par with any of the commercial fraud detectors as long as that fraud detector’s results are included in the system’s information when it creates its cases. The advantage of this fraud case builder is that it should be able to work faster than human auditors building these cases. Attempting to replicate the human auditor’s task is what is unique to this system in contrast to fraud detection systems that currently exist which function primarily as tools for the human auditor to identify records and situations that the auditor must still investigate.

As a comparison, the authors needed to compare their performance not to existing fraud detectors, since the proposed system would actually use those fraud detector’s results, but rather to the performance of human auditors in identifying actual cases of fraud. The proposed system should either be able to find all of the cases that the human auditors find, but faster, or cases that were not known to the human auditors.

Three experiments were conducted:

1. A preliminary experiment demonstrating the utility of combining information from more than one fraud detector to produce more correctly identified cases of fraud over the fraud cases of two separate detectors each working alone.
2. An experiment to measure the effects of combining information from two different fraud detectors in varying ratios.
3. A performance comparison on processing times of the proposed system against actual human fraud-auditor-created rule sets on real health insurance data.

Background

As noted by Li et al.[3], health-care-fraud behaviors may be classified under the potential involved parties:

- i. Service provider fraud
- ii. Insurance subscriber fraud
- iii. Insurance carrier fraud

Each of these parties may commit fraudulent actions such as

- i. Service-provider fraud: Billing for services not actually performed
- ii. Insurance-subscriber fraud: Falsifying claims to medical services never received
- iii. Insurance-carrier fraud: Falsifying benefits statements

In the case of a false billing of a service never performed, this could be perpetrated by either the provider or subscriber and illustrates the importance of relating a billing anomaly with is the person ultimately benefiting from the false bill in order to correctly identify the guilty and avoid false accusations.

Uncovering such relations may require information from multiple resources such as the provider’s billing records compared against the insurance subscriber’s health records. Such records can be, and often are, in different formats, requiring a system flexible in searching and matching across these multiple formats. It is this required flexibility in representation combined with a capability to search and build patterns of relations that the authors chose a semi-supervised learning technique known as reinforcement learning.

Reinforcement Learning

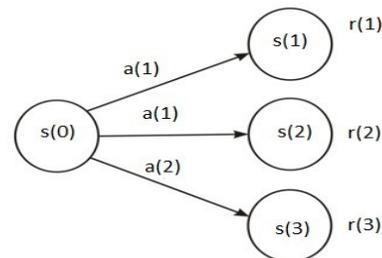


Figure 1. Representation of a Reinforcement Learning Network

Here, the authors provide a very brief overview of the reinforcement learning technique to machine learning. For a more detailed explanation, see Sutton and Barto [4]. In reinforcement learning, an environment is modeled as a network of states, S . Each state, $s \in S$, is associated with a set of possible actions, $a(s) \in A$ and a reward for entering that state $r(s) \in R$, where A and R are sets of rewards and actions, respectively. It is possible to transition from one state, $s(i)$, to another, $s(j)$, by choosing an action, $a(s(i))$, with a certain probability, $\text{Prob}(s(j) | s(i), a(s(i)))$. The advantage of this representation is that it places very little restriction on formatting. The objective of this representation is to find an optimal policy. A policy is a function that maps states to

actions. In other words, it makes choices on actions to take for any given state visited. An optimal policy maximizes the long-term rewards one may obtain as one navigates through the network [4-6]. For the proposed fraud-auditing system, the optimal policy is one which will maximize the likelihood that a set of related data records form a case of fraud. This relation is made by allowing rewards, R, in the reinforcement environment to represent a numerical value derived from fraud detectors, where the larger the reward value, the more likely attributes in the state are an indicator of fraud. The appropriateness to use reinforcement learning, a method traditionally used in robotic search, to search across database tables was presented by Lu [7].

Fraud Detectors

As stated in the introduction, a variety of fraud detection methods exist, with the adaptive ones being divided into supervised and unsupervised methods [8], [9]. For supervised methods, both fraudulent and non-fraudulent records are used to train a system, which then searches and classifies new records according to the trained patterns. The limitation of supervised methods is that one must have both classes of records identified for the system to train on. Thus, this approach is limited to only previously known methods of committing fraud. A few of the more popular supervised approaches involve Bayesian networks [10] and classifiers to detect suspicious claims [11], Neural networks [12], [13] and Decision trees [14], [15].

Unsupervised methods, in contrast, often identify records that do not fit expected norms or essentially looking for outliers [16]. The advantage of this approach is that one may identify new instances of fraud. A common approach to this method is to use forms of outlier detection. The main limit to this approach is that we are essentially identifying anomalies that may or may not be fraudulent behavior. An audit investigator in this case may then be employed to analyze these anomalies for their likelihood to be indicative of fraud. One expert system that has been developed using unsupervised techniques is known as SmartSifter [17], which uses probabilistic models to generate its outliers.

Application Method

Reinforcement learning is well-suited to linking together states through its state-action policy mapping. For reinforcement learning to be used as a fraud case builder, it needs to be able to relate rewards with outliers that are indicative of possible fraud. It does so by preprocessing all records using sets of fraud detectors such as outlier/anomaly methods, so that all records have a numerical reward value that represents their likelihood of fraud. The states of the proposed RL environment relate to individual records of the application environment, and the actions are the attributes of

a record. In this way, two records with the same attributes are linked together by a common attribute just as an action can relate two states of a classic reinforcement learning environment network. A second phase of the preprocessing is to search record columns for matching attributes across different tables so that the columns may be linked during the reinforcement learning search. For example, two databases may both include a column for patient names. These should be linked in order to match records across databases.

After the preprocessing, an exploration search on the RL environment produces an optimal policy indicative of fraud. This policy will then be followed to uncover a list of records that are related with a high chance of fraud. Following is a summary of the proposed approach.

1. First, preprocess all database records with a selection of fraud detectors to assign reward values.
2. A second preprocess phase links database columns through their attributes matching different database tables using a pattern match search on the values within the columns.
3. Run a reinforcement learning approach using attributes of a record as action choices in a reinforcement learning context to search the databases until an optimal policy is found.
4. Navigate through the environment using the derived optimal policy with a start state drawn from one of the significant outliers from the statistical distributions produced from step 1.
5. Return all records encountered.

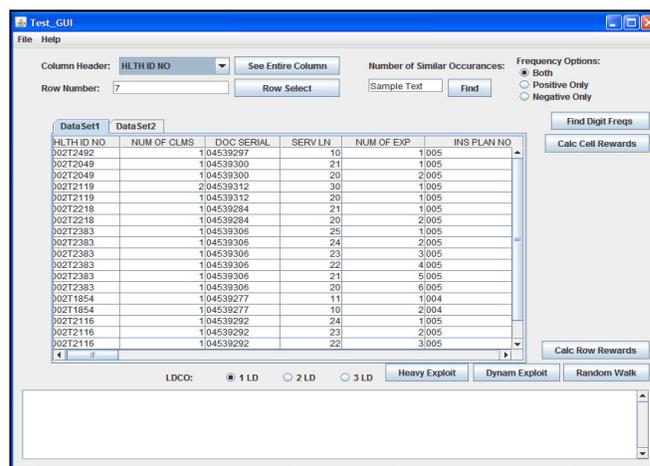


Figure 2. Screen capture of the application with sample health insurance data set

The returned records are all in ranked order from most likely fraudulent to least likely, based on their reward value. For further details on these algorithmic techniques and principles behind this application see Lu [7].

Figure 2 illustrates the interface window of the authors' software application. There are a number of features that have been incorporated for usability and functionality. To allow for searching multiple database tables, the authors used multiple tabs; each tab holds a table of a database. Dataset 1 in Figure 2 illustrates a sample health insurance record, while data set 2 shown in Figure 3 illustrates a table of hospital in-take records. The blank window below the data-sets window is where the list of related records that are potentially fraudulent are presented, with each potential case of related records listed per row. The system has capabilities to search for specific terms within columns of tables in order to support a user's analysis of the results. This initial prototype includes functionality for:

1. Two forms of fraud outlier detectors;
2. Search capability to set thresholds for different random walks;
3. Options on using an online update method with the 'Dynamic Exploit' button, which continuously updates the policy with new updated records or an offline solution with the 'Heavy Exploit' button.

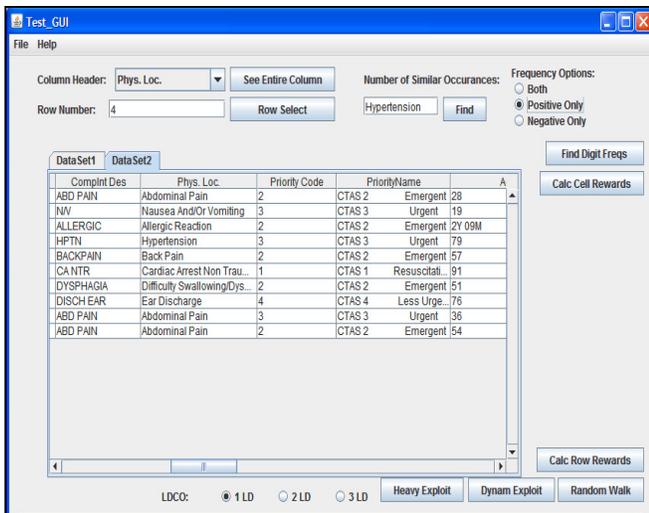


Figure 3. Screen capture of application with sample hospital intake records displayed

Experiments

Experiment 1

As a first test on the proposed automated auditing approach, the authors ran a system test using real insurance data consisting of 31,804 records that had been audited and labeled for fraud cases. The data was divided into a set of ten test sets with a random selection method to choose the records that would be included in each test case. An outlier detection approach, where records that deviated from a set threshold were included in the fraud category, were com-

pared against the proposed system which used the same threshold but with aggregated information over several related records. A single fraud detector is used in any given record, but different records could use different fraud detectors. In the current experiment, the authors used two fraud outlier detectors: a Normal Distribution outlier and a method known as Benford's law distribution outlier.

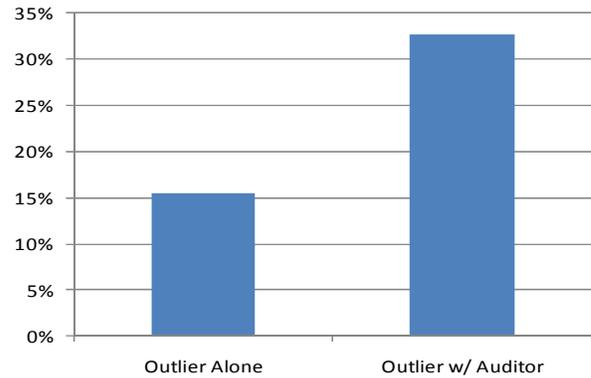


Figure 4. Improvement with automated auditor

Figure 4 illustrates the averaged results of this experiment comparing the percentage of correctly identified fraudulent cases relative to the number of cases each method recommended as potentially fraudulent. Generally, the automated auditor had approximately double the precision of using the fraud detector alone.

Experiment 2

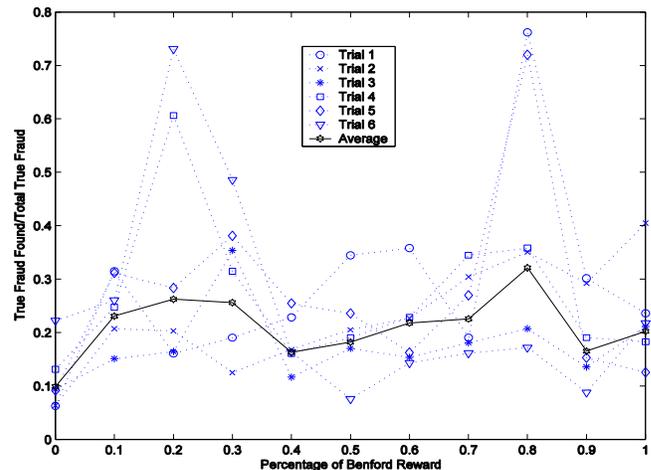


Figure 5. Combining fraud detection data with Benford vs. Normal Distribution Rewards

As a second test, the authors wanted to consider the scenario where more than one fraud detector may contribute useful fraud information to a record. Under such a situation, how should one combine the multiple pieces of evidence into a single reward value? In this experiment, the authors

combined normal distribution outlier rewards with the Benford's Law outlier reward. The reward formula used was

$$\text{Reward}(i) = \text{RB}(i) \times \%B + \text{RN}(i) \times (1 - \%B), \quad (1)$$

where $\text{RB}(i)$ is the Benford reward calculated for state i , $\text{RN}(i)$ is the Normal distribution reward calculated for state i and $\%B$ is the percentage Benford's reward contributes to the overall reward value ($0 \leq \%B \leq 1$).

Using real industry data that had been audited for fraud, a reinforcement learning algorithm was run over six trial runs, where each trial was comprised of fifteen sampling trajectories of 1000 sampling steps and where rewards were varied in increments of 10%. Normal distribution rewards appear on the far left side of Figure 5 and decrease from there with increasing amounts of Benford rewards until 100% Benford's Law distribution rewards on the far right side of the figure are reached. The data used includes 227,156 records with 1,526 previously identified fraudulent records, which were used to verify the accuracy of the fraud results.

Analyzing the average results from the results of Figure 5, it would also appear that the two reward methods interact in a complicated manner. The fraud-detection accuracy increases to a peak at 20% Benford/80%Normal, and then peaks again at approximately 80% Benford/20% Normal, giving a possible bimodal interpretation. This indicates that combining the two reward mechanisms has some benefit yielding generally greater true-fraud cases compared to the detection accuracies of each detector alone shown on each far end of the graph. This result supports the idea that combining information from multiple fraud sources will generally yield improved results.

Experiment 3

In the third experiment, an insurance business dealing in coverage for individuals for both dental and pharmaceutical drug claims wished to assess their auditors' performances against the proposed systems to determine how much efficiency improvement, if any, there may be using this automated fraud auditor. The experiment looked at comparing the performance time of a human-auditor-created-fraud rule using customized commercial software tools against the proposed automated fraud auditor. The rule sets were designed to extract records satisfying conditions that they had identified as indicative of potential fraud. The rules also included a rank ordering of the records from most-likely-fraudulent to least. The company used an oracle database with customized query capability with commercial statistical software for the fraud auditors to extract their records. A base requirement was that the ranking should be the same on the same set of data. If that condition was satisfied then the performance times of the human-auditor-created rule set

query and ranking time could be compared with the authors' system's time.

The data consisted of three months of drug claims records comprised of 1,890,548 actual records from January 1, 2010 to March 31, 2010. The authors were asked to conduct two searches: the first was a 'Field test' of the proposed system. The second was an 'Unknown results test' that was actually a search that the business' own fraud system failed to complete. During a test, their system simply crashed during the query processing; thus, they were unaware of what actual results they should get. The authors' system was able to produce the required rank ordering for the first test and completed the second search with a list of rankings. The runtimes are presented in Table 1. The table compares the actual search and ranking times of the proposed system against the time the insurance business' rule set queries took to conduct the same search.

Table 1. Runtime Comparisons of Fraud Claims with ranking order

	Automated Fraud Auditor	Business' Fraud Auditor's system using rule sets
Field test	18 min 22 sec.	Approximately 7 hours
Unknown Results test	4min. 18sec.	Failed to Complete

The results of this experiment demonstrated that not only did the system produce results satisfying the required criteria, but that the runtimes were impressively better. This supports the premise for the creation of this system to support human auditors by replicating their task. But by automating portions of the task, the human auditors can be allowed to conduct larger and more diverse searches with the saved time.

Conclusions

In this paper, the authors presented a case study of software implementation for an automated fraud auditor designed to support human auditors in their search for fraudulent acts. The main advantages of this approach are its flexibility in combining information from multiple database records and using the information from multiple fraud detectors to build its cases. A detailed approach was used to allow for this flexibility by using a semi-supervised learning approach with few restrictions on format or structure. Preliminary tests on the prototype were conducted incorporating two outlier fraud-detection methods and demonstrated its improved precision for finding fraudulent records over the outlier fraud detectors alone. Different approaches were tested for combining fraud detectors. Finally, runtime improvements were demonstrated over human-auditor-created

fraud searches on real health insurance data producing comparable results but in a significantly shorter time frame.

In terms of future work, to test the approach's efficacy, the authors plan to incorporate more sophisticated fraud detectors to build fraud rewards such as a Bayesian [11] and a neural network detector [18]. The ultimate objective is to measurably recover losses due to uncovered cases of actual fraud. However, issues of privacy [2], [19], maintaining corporate and institutional confidence [1], and the speed of actual prosecution of potential fraud cases [20] presents challenges for both the measuring of actual cost savings and publicly presenting those savings.

Acknowledgements

This work was supported by The Ontario Partnership for Innovation and Commercialization and the Natural Sciences and Engineering Research Council of Canada.

References

- [1] Stead WW, Lorenzi NM. Pryor. Health informatics: linking investment to value. *JAMIA*. 2010;6:341-348.
- [2] Simborg DW. Healthcare fraud: whose problem is it anyway? *JAMIA*. 2008;15:278-280.
- [3] Li J, Kuei-Ying Huang K-Y, Jin J, Shi J. A survey on statistical methods of health care fraud detection. *Health Care Management Sci*. 2008;11:275-287.
- [4] Sutton RS, Barto AG. Reinforcement learning: an introduction. MIT Press. 1998.
- [5] Sutton RS. Learning to predict by the method of temporal differences. *Machine Learning*. 1988;3:9-44.
- [6] Singh SP, Sutton RS. Reinforcement learning with replacing traces. *Machine Learning*. 1996;22:123-158.
- [7] Lu F. Uncovering fraud in direct marketing data with a fraud auditing case builder. *PKDD*. 2007;11:540-547.
- [8] Fawcett T. AI Approaches to fraud detection and risk management. *AAAI Workshop: Technical Report*. 1997;WS-97-07.
- [9] Bolton RJ, Hand DJ. Statistical fraud detection: a review. 1999;17(3):235-255.
- [10] Ormerod T, Morley N, Ball L, Langley C, Spenser C. Using ethnography to design a mass detection tool (MDT) for the early discovery of insurance fraud. *Proceedings of ACM CHI*. 2003.
- [11] Chan CL, Lan CH. A data mining technique combining fuzzy sets theory and Bayesian classifier-an application of auditing the health insurance fee. *Proceedings of the Inter. Conf. on A.I.* 2001;402-408.
- [12] Hall C. Intelligent data mining at IBM: new products and applications. *Intelligent Software Strategies*. 1996;7(5):1-11.
- [13] Ortega PA, Figueroa CJ, Ruz GA. A medical claim fraud/abuse detection system based on data mining: a case study in Chile. *Proceedings of Int. Conf. on Data Mining*. 2006.
- [14] Bonchi F, Giannotti F, Mainetto G, Pedreshi D. A classification-based methodology for planning auditing strategies in fraud detection. *Proceedings of SIGKDD99*.1999:175-184.
- [15] Williams G, Huang Z. Mining the knowledge mine: The Hot Spots methodology for mining large real world databases. *Lecture Notes in Computer Science*. 1997;1342:340-348.
- [16] Langley P. Elements of machine learning. Morgan Kaufmann, San Francisco. 1995.
- [17] Yamanishi K, Takeuchi J, Williams G, Milne P. Online unsupervised outlier detection using finite mixtures with discounting learning algorithms. *Data Mining & Knowledge Discovery*. 2004;8:275-300.
- [18] He H, Wang J, Graco W, Hawkins S. Application of neural networks to detection of medical fraud. *Expert Syst Appl*. 1997;13:329-336.
- [19] Buckovich SA, Helga ER, Rozen MJ. Driving toward guiding principles: a goal for privacy, confidentiality, and security of health information. *JAMIA*. 1999;6:122-133.
- [20] National Health Care Anti-Fraud Association website. Available at: <http://www.nhcaa.org>. Accessed March 2010.

Biography

FLETCHER LU is an Assistant Professor in the Faculty of Health Sciences at the University of Ontario Institute of Technology. He is peer-review published in areas of health informatics, machine learning and artificial intelligence. He received a Bachelors of Mathematics with distinction as well as a PhD. in Computer Science from the University of Waterloo. Professor Lu may be reached at fletcher.lu@uoit.ca

A NEURAL OSCILLATOR MODEL FOR TINNITUS AND ITS MANAGEMENT BY SOUND THERAPY

Hirofumi Nagashino, The University of Tokushima; Ken'ichi Fujimoto, The University of Tokushima; Yohsuke Kinouchi, The University of Tokushima; Ali A. Danesh, Florida Atlantic University; Abhijit S. Pandya, Florida Atlantic University

Abstract

Tinnitus is the perception of sound occurring without an external stimulus. Sound therapy is one of the most effective treatment techniques that have been employed by clinicians. To account for mechanisms of tinnitus generation and the clinical effects of sound therapy from the viewpoint of neural engineering, the authors have proposed a conceptual neural oscillator model with plasticity for the human auditory system. In this study, the authors found that this model has a bistable state, i.e., a condition where a stable oscillatory state and a stable non-oscillatory equilibrium state coexist at a certain parameter region. It was also found that the oscillation can be inhibited by supplying sinusoidal or random stimuli, which was hypothesized as sound for treatment of tinnitus, in this model. Through numerical simulations, the authors discovered that adequate noise stimulus can inhibit the oscillation. By hypothesizing that the oscillation and the equilibrium correspond to generation and inhibition of tinnitus, respectively, it can be stated that these phenomena could explain the fact that the human auditory system temporarily halts perception of tinnitus following sound therapy. This paper describes dynamical properties of the model and inhibition of the oscillation for sinusoidal stimulus and a kind of noise stimulus, which corresponds to the sound in the treatment of tinnitus.

Introduction

Tinnitus is the perception of phantom sounds in the ears or in the head. It is believed that tinnitus is an auditory phantom phenomenon resulting from neuronal activity somewhere along the auditory pathway. This phantom perception is a common condition and it can originate from many sources. Tinnitus can be perceived due to the damages in a variety of the pathologies of the auditory system. It has been shown that auditory percept can be generated by exposure to loud noise or medications which have a toxic effect on the inner ear. Additionally, tinnitus can be associated with a variety of diseases such as thyroid abnormalities, diabetes and hypertension. In some cases, tinnitus can be traced to an internally generated sound, for example, spontaneous otoacoustic emissions. However, in the overwhelming majority of serious sufferers, there is no obvious sound source to account for the tinnitus percept [1-3].

Tinnitus is a chronic disease with a reported prevalence of 10–15% [4–8]. Globally speaking, this corresponds to ap-

proximately 700 million people. The effect of tinnitus on quality of life in patients suffering from this disease should not be underestimated. It is noteworthy that an estimated 20% of patients indicate that their quality of life is significantly deteriorated. Many patients experience insomnia and depression, and in 1% of the population tinnitus seriously interferes with their life [7–9].

Attempts have been made to help those who suffer from tinnitus. The common types of tinnitus treatment include medication therapy, biofeedback, relaxation therapy, cognitive behavioral therapy and sound therapy. Although many therapies have been proposed and tried, there is no systematic and proven approach for treating tinnitus.

Despite numerous animal and human studies, the neural abnormalities underlying tinnitus are largely unknown. In some patients tinnitus can be traced to an internally generated sound such as vascular structures in pulsatile tinnitus; in the vast majority, however, no obvious sound source can be pinpointed. It has been proposed that tinnitus results from abnormal neuronal activity arising at some point along the auditory pathways which is interpreted as sound at a cortical level [10–12]. This abnormal neuronal activity is hypothesized to be the neural correlate of tinnitus, which is considered to be an auditory phantom phenomenon, similar to central neuropathic pain, due to neural plasticity in response to total or partial deafferentation somewhere along the auditory tract [10], [13–15]. Animal and human studies have provided some evidence for this theory [8], [12], [13], [16–22]. Functional MRI has been applied in a few studies, mainly case studies [23–25]. The purpose of these studies has been to visualize the entire central auditory pathway in patients with lateralized tinnitus and in patients with bilateral tinnitus to evaluate lateralization of fMRI signal change and activation clusters. The most important fMRI study shows an abnormally low percentage signal change in the inferior colliculus (IC) contralateral to the side of the perceived tinnitus [26]. Structural brain changes in tinnitus have also been discovered using MRI [27].

The role of neural plasticity in the auditory system and tinnitus has been discussed in neurophysiologic studies, and thalamocortical correlates or dorsal cochlear nucleus activities with plasticity have been investigated [28-32]. Electrophysiological studies of the auditory system have demonstrated an evidence for thalamic plasticity via top-down modulation [29]. Computational modeling of thalamocortical correlates with plasticity from the perspective toward un-

derstanding of the tinnitus has been reported [33]. A tinnitus model based on the work by Jastreboff [11], combined with the adaptive resonance theory of cognitive sensory processing [34], has been proposed for identification of neural correlates of the tinnitus decompensation [35]. The effect of auditory selective attention on the tinnitus decompensation has also been investigated by modeling corticothalamic feedback dynamics [36], [37].

There are two typical sound therapy techniques, namely, the TM (Tinnitus Masking) technique and TRT (Tinnitus Retraining Therapy), where those who suffer from tinnitus listen to these therapeutic sounds for several hours a day [38]. In these techniques, white noise or spectrum-modified white noise are introduced to tinnitus sufferers as therapeutic sound. These sounds are usually presented via a custom-made noise (sound) generator or a tinnitus masker. It has been reported that sound therapy has a clinical effect, in a great number of cases in this management method where tinnitus perception is temporarily halted after the removal of the noise (sound) generator. This cessation of tinnitus following the presentation of a masking stimulus is referred to as residual inhibition. The mechanisms of tinnitus and its management by sound therapy, however, are not clear. Some attribute the success with sound therapy to brain plasticity [39] while others consider it a habituation process [40].

The purpose of this current study was to address mechanisms of tinnitus generation and the clinical effect of the sound therapies from the viewpoint of neural engineering. Accordingly, the authors proposed a plastic neural network model for the human auditory system. They also previously reported [41] that a certain region of the parameter hyper space exists where an oscillatory state and a non-oscillatory equilibrium state coexist. It was shown that the oscillatory state is inhibited by supplying a sinusoidal stimulus resulting in a transition to an equilibrium state [42], [43]. By hypothesizing that the oscillation and the equilibrium correspond to generation and inhibition of tinnitus, respectively, the authors demonstrated that these phenomena could explain the fact that the human auditory system temporarily halts perception of tinnitus following sound therapy. This paper illustrates inhibition of the oscillation in the proposed model using band-noise stimulus as therapeutic sound [44], [45]. In tinnitus clinics across the globe, similar noise stimuli have been employed for treatment of tinnitus by TM.

Model Description

The human auditory system consists of two divisions, a peripheral portion and a central portion. The cochlear hair cells are located in the peripheral portion and transform acoustic vibrations received by the ear into neural signals. The central auditory pathway is composed of many portions.

It receives the neural auditory messages that have travelled via the auditory nerves to the cochlear nucleus, the superior olivary complex, the inferior colliculus, the medial geniculate body, and the auditory cortex. Subsequently, the brain perceives the neural signals as sound. In cases of certain tinnitus patients it has been observed that cochlear dysfunction occurs, and abnormal neural signals from the cochlea cause abnormality in the central nervous system. Consequently, tinnitus can be triggered. In addition to ascending pathway, the cochlear nucleus complex receives descending efferent fiber bundles to control the function of cochlear outer hair cells [9].

In this study, the authors proposed a conceptual neural network model to account for tinnitus generation and its inhibition from the neural engineering point of view [33]. Figure 1 shows the proposed neural oscillator model of the human auditory system.

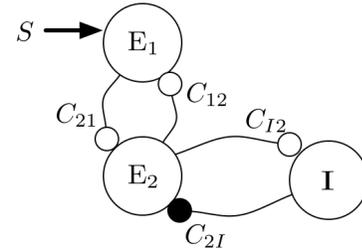


Figure 1. A Neural Oscillator Model

The human auditory system in this case is represented as a neural oscillator which consists of two excitatory units denoted by “ E_1 ” and “ E_2 ”, and an inhibitory unit denoted by “ I ”. The unit represents the aggregate of a neural ensemble in this model. The excitatory units E_1 and E_2 form a positive feedback loop by mutual coupling, while the units E_2 and I form a negative feedback loop by mutual coupling. These two loops enable the model to oscillate. This configuration is the simplest in terms of neural arrangement that could demonstrate oscillatory behavior. The unit E_1 receives an incoming signal, S , which is associated with an external sound signal.

The neural coupling from the j -th unit to the i -th unit is expressed by the positive constant C_{ij} ($i, j \in \{1, 2, I\}$).

$$\frac{dx_1}{dt} = (-x_1 + C_{12}z_2 + S) / \tau_1 \quad (1)$$

$$\frac{dx_2}{dt} = (-x_2 + C_{21}z_1 - C_{2I}z_I) / \tau_2 \quad (2)$$

and

$$\frac{dx_j}{dt} = (-x_j + C_{12}z_2)/\tau_j \quad (3)$$

where x_j and τ_j are the internal potential and time constant of the j -th unit, respectively. The output of the j -th unit is denoted by z_j , which is given by the equation

$$z_j = \frac{2}{\pi} \tan^{-1} x_j \quad (4)$$

It was assumed that the coupling strength from the unit E_2 to the unit E_1 , denoted by C_{12} , would have plasticity in such a way that it would change according to the product of the outputs of the units E_1 and E_2 . This would mean that the coupling strength C_{12} would be one of the state variables in the model system. It was expressed as

$$\frac{dC_{12}}{dt} = (-C_{12} + bz_1z_2 + C_0)/\tau_c \quad (5)$$

The C_0 , b , and τ_c are also positive constants which denote the equilibrium of C_{12} under $z_1z_2 = 0$, the efficiency of strengthening the synaptic coupling based on Hebbian hypothesis [46], and the time constant of C_{12} , respectively.

At the present time, it is not possible to specify what regions in the brain correspond to each unit in the model. The model was hypothesized to represent tonotopic organization and depends on the perceived pitch and reported frequency of tinnitus. Based on the anatomical structure of the auditory system, the proposed model is likely to include the thalamus, at which a massive corticofugal projection ends. The thalamo-cortico-thalamic loop forms an ideal positive oscillatory loop, while the thalamic interneurons and thalamic reticular GABAergic neurons likely play the role of inhibitory neurons.

The external auditory stimulus, which is represented by S in Figure 1, is received as an input in unit E_1 and results in generation of aggregate neuronal activity of an ensemble in the proposed model. In the auditory system, such processing occurs at the peripheral nervous system and the corresponding mechanism in terms of neural engineering is represented within unit E_1 . Aggregate activity of thalamic interneurons and thalamic reticular GABAergic neurons are captured within the excitatory unit E_1 and inhibitory unit I .

Aggregate neuronal mechanism represented within the cortex pertaining to the perception of tinnitus is represented by the unit E_2 . The thalamo-cortico-thalamic loop is represented by the excitatory links between the units E_1 and E_2 and the excitatory-inhibitory links between units E_2 and I .

Attractive Region of Non-Oscillation

The plastic system expressed by equations (1) – (5) has two attractors. An equilibrium point in the dynamical system can easily be found: $(x_1, x_2, x_j, C_{12}) = (0, 0, 0, C_0)$. The other attractor is an oscillatory orbit. Numerical analysis of the system revealed that the equilibrium exists in the range of C_0 , $0 \leq C_0 \leq 8.06$ and the oscillatory orbit in $C_0 \geq 2.65$ [40]. Consequently, the system is bistable in $2.65 \leq C_0 \leq 8.06$.

Also in the system with no plasticity—that is, where the coupling strength C_{12} is not expressed by equation (5), but rather given to a constant value—there are two attractors: an equilibrium point $(x_1, x_2, x_j) = (0, 0, 0)$ and an oscillatory orbit. The system also has the similar bistable region of C_{12} to the region of C_0 in the plastic system. Which attractor the system converges to depends on the initial values of x_1, x_2 and x_j .

In Figure 2, are shown two examples of the region of the initial values of x_1 and x_2 . When coupling strength values C_{12}, C_{21}, C_{2j} and C_{j2} are help constant, the system is attracted to the non-oscillatory solution when different initial values of x_1 and x_2 are used. The parameters in equations (1) – (3) were fixed such that $\tau_1 = 0.01$ seconds, $\tau_2 = 0.01$ seconds, $\tau_j = 0.02$ seconds, $C_{2j} = 10$, $C_{j2} = 10$ and $C_{j2} = 20$. The initial value of x_j was also fixed at zero.

The dynamics of the system are governed by equations (1) – (3). In the black region in Figure 2, if the initial x_1, x_2 values are inside the black region—excluding $(0, 0)$ —then the system is attracted to the equilibrium point $(0, 0, 0)$ and the oscillations subside. If the initial x_1, x_2 values are outside the black region, the oscillation occurs as the system is attracted to the oscillatory orbit.

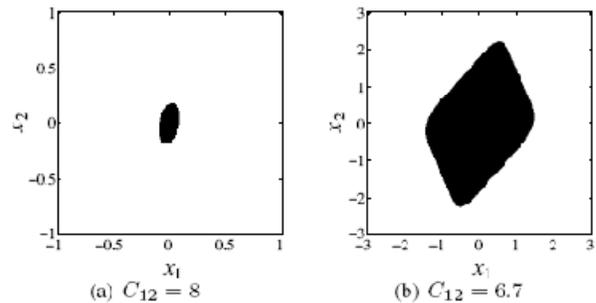


Figure 2. Regions of attraction to the equilibrium point at all constant couplings, i.e., all C_{ij} are constant. The initial value of x_j is fixed at zero

Notice that these results show dynamic properties of the non-plastic neural oscillator model, which is described by equations (1) - (3). According to the results, the model has the dynamic property that the attractive region of the non-oscillatory solution is expanded by reducing the value of C_{12} . In addition, any oscillatory behavior in this region eventually settles down in the non-oscillatory state without external stimulus according to the dynamic property of the model. Eventually, when $C_{12} < 2.65$, any initial value yields a non-oscillatory solution. Therefore, to inhibit the oscillation, it is important that the states of x_1, x_2, x_I, C_{12} change into such an attractive region of non-oscillation by external stimulus.

Inhibition of Oscillation by External Stimuli

The inhibition of oscillation by various external stimuli was examined with the model incorporating the synaptic plasticity, which is the system described by equations (1) - (5). In equation (5), the plasticity parameter values $C_0=3, b=20, \tau_c=0.5$ seconds were employed. These parameters of plasticity were arbitrarily determined so that the simulation was performed within the appropriate time. The time scale would be much longer in the clinical situation as it is important to sufficiently expose the ears to a sufficient duration of acoustic stimulation for better inhibition or habituation.

Demonstrated here are some simulation results with sinusoidal wave and band noise as external stimuli. The stimuli were applied to the unit E_1 for the duration of $2 \leq t \leq 8$ seconds.

1) Sinusoidal stimulus

Auditory stimulations can be viewed as a composite of sinusoidal signals. Hence, the first experiment was supplied an adequate sinusoidal stimulus [28]. Figure 3 shows an instance in which the oscillation was inhibited by sinusoidal stimulus defined as

$$S = 2\sqrt{2} \sin 20\pi t. \quad (6)$$

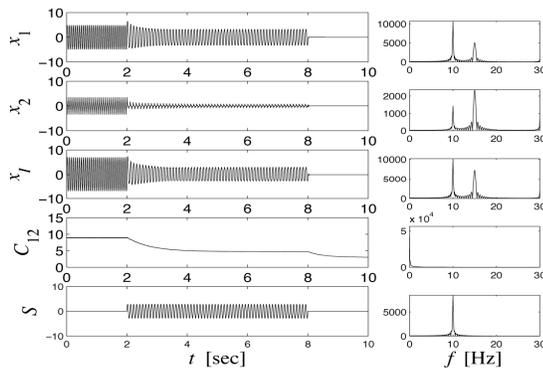


Figure 3. Inhibition of oscillation by sinusoidal stimulus

The value of the amplitude was fixed so that the root mean squared value (RMS) of the stimulus was 2.0. The timing diagrams in the left column show the waveforms of x_1, x_2, x_I, C_{12} and S , respectively; and their corresponding power spectra are illustrated in the right column.

It can be seen that by applying stimulus S starting at 2 seconds, the value of C_{12} gradually decreases and, consequently, the oscillations are inhibited. Note that the oscillation does not reoccur even after the sinusoidal stimulus stops at 8 seconds. By hypothesizing that the damping of C_{12} oscillations corresponds to what is called residual inhibition in the human auditory system for tinnitus, the phenomenon could explain the fact that the human auditory system temporarily inhibits the perception of tinnitus after sound therapy. This is a promising observation and it can help in better management of tinnitus.

The reason why the coupling strength C_{12} decreases is explained as follows. The plasticity is formulated based on the Hebbian hypothesis in equation (5). It works in such a way that the coupling strength C_{12} increases when the units E_1 and E_2 oscillate in-phase, and decreases when the oscillations of the units E_1 and E_2 are anti-phase. Without stimulus, the oscillations of the units are in-phase, while they are out-of-phase with stimulus and close to anti-phase.

2) Band-noise stimulus:

Band-noise stimulus is typically used in the treatment of tinnitus using the TM approach [27]. In this approach, the desired noise is a band of noise with a frequency emphasis that approximates the frequency of perceived tinnitus. The frequency (pitch match) of tinnitus can range from low frequencies to high frequencies in different individuals. Most of the tinnitus sufferers perceive tinnitus at high frequencies between 2–8kHz.

In this experiment, the authors hypothesized that the fundamental frequencies of perceived tinnitus were 2, 4, 6, and 8kHz, and subsequently employed a band noise generated from Gaussian white noise through a band-pass filter which operates between each fundamental frequency with a $\pm 5\%$ margin. RMS (Root Mean Square) was also adjusted to about 400.

Figures 4(a)-(d) show successful results for inhibition of the oscillations. The timing diagrams in the left column show the waveforms of x_1, x_2, x_I, C_{12} and S , from top to bottom, respectfully. Their corresponding power spectra are illustrated in the right column.

In Figures 4(a) and 4(b), oscillation of x_2 and x_I stops immediately after the noise input is applied and it brings the rapid decrease of C_{12} . In Figure 4(c), oscillation of x_2 and x_I is maintained for about 1 second. In Figure 4(d), oscillations

of x_2 and x_I are maintained almost until $t=8$ seconds. However, the value of C_{12} gradually decreases from the point of applying the stimulus. Eventually, the oscillations of x_2 and x_I stop. This implies that plasticity in connectivity between various neuronal ensembles in the auditory system may play a role in the inhibition of tinnitus. Note that during the application of noise, input unit E_1 is in an oscillating mode replicating the perception of tinnitus. After removing the input, the oscillation of the unit is in the state of non-oscillation, which replicates the state where the tinnitus is not perceived. It was observed that the oscillation was inhibited with certainty in all the simulations when 100 trials were conducted with different random sequences.

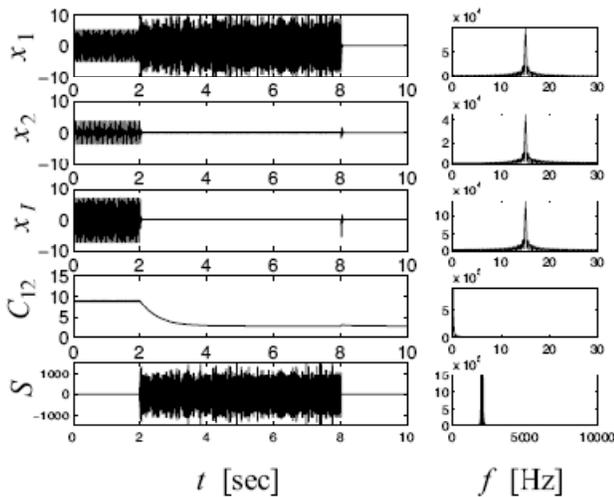


Figure 4(a). Inhibition of oscillation by band noise stimulus with the band between 2kHz \pm 5% margin and 400 RMS

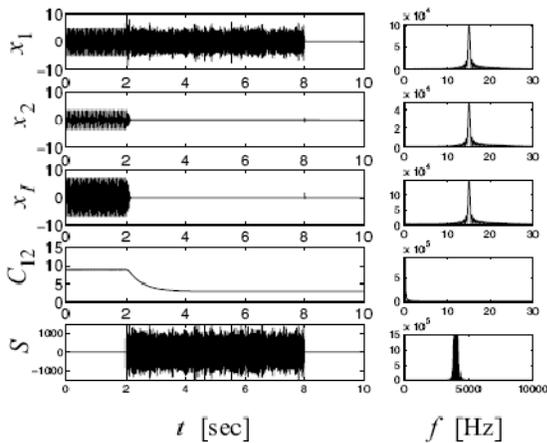


Figure 4(b). Inhibition of oscillation by band noise stimulus with the band between 4kHz \pm 5% margin and 400 RMS

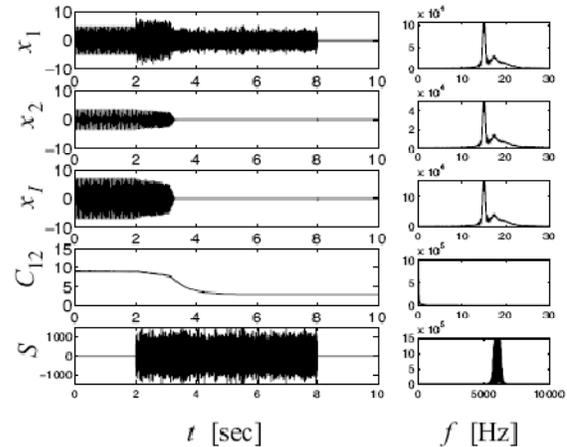


Figure 4(c). Inhibition of oscillation by band noise stimulus with the band between 6kHz \pm 5% margin and 400 RMS

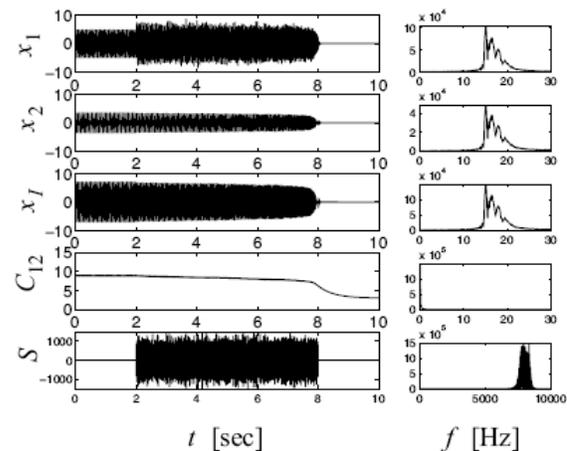


Figure 4(d). Inhibition of oscillation by band noise stimulus with the band between 8kHz \pm 5% margin and 400 RMS

Figures 5(a) and 5(b) show the results for a 4kHz band-pass filter with a margin of \pm 5% and an RMS adjusted to about 200. Figure 5(a) shows a successful result similar to that of Figure 4(d). Oscillations of x_2 and x_I were maintained almost until $t=8$ seconds. However, the value of C_{12} gradually decreased from the point of applying the stimulus. Eventually, the oscillations of x_2 and x_I stopped. On the other hand, for the same conditions, there were unsuccessful results in which the value of C_{12} was almost unchanged for the 4kHz filter and, consequently, the oscillations continued through the entire simulation time, as seen in Figure 5(b). It was observed that the oscillation was inhibited with the probability 74% when 100 different random trials were conducted.

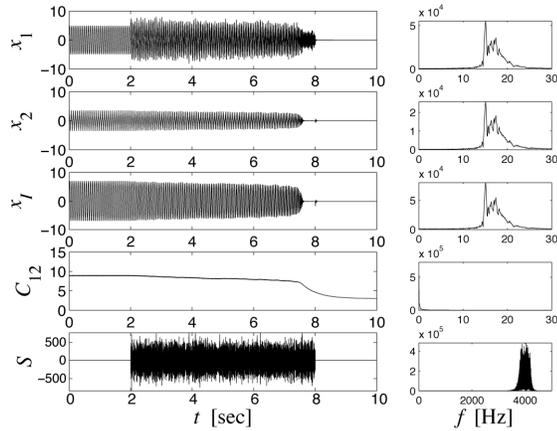


Figure 5(a). A successful experiment for inhibition of oscillation by band noise stimulus with the band between 4kHz \pm 5% margin and 200 RMS

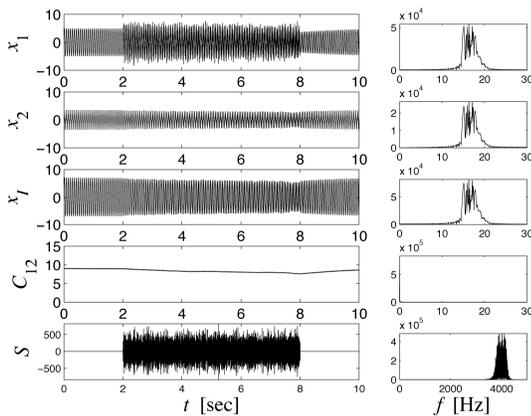


Figure 5(b). An unsuccessful experiment for inhibition of oscillation by band noise stimulus with the band between 4kHz \pm 5% margin and 200 RMS

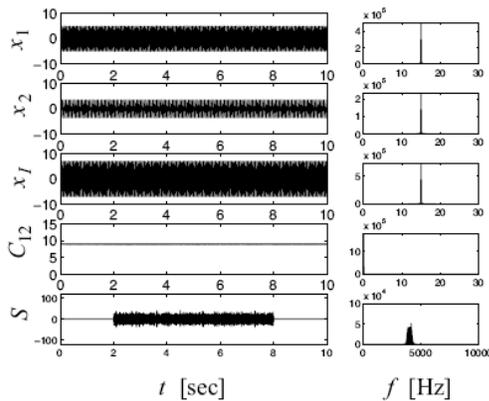


Figure 6. An unsuccessful experiment for inhibition of oscillation by band noise stimulus with the band between 4kHz \pm 5% margin and 10 RMS

As seen in Figure 6, for band noise with much lower power (RMS=10), it was observed that the value of C_{12} was almost unchanged and, consequently, the oscillations continued through the entire simulation time for all the simulations that we examined with 100 different random sequences. Therefore, it can be concluded that the inhibition of the oscillations by band noise stimulus requires an appropriately higher RMS value.

For TM that uses band noise, higher amplitudes were used for the external sound in clinical practice so that the patients could not hear the sound of tinnitus [27]. It was not possible to precisely compare the RMS values of the noise in the simulation with the clinical data. However, the amplitude of the oscillation of x_1 was larger than the one that was seen before the noise was applied in Figures 4 and 5(a), and successful results were obtained. In Figure 6, the amplitude of the oscillation of x_1 was same as the one that was seen before the noise was applied and the result was unsuccessful. It corresponded to the situations when the noise was applied in clinic. Therefore, the results seem to be consistent with the practice.

Conclusions

In this study inhibition of the oscillation in the plastic neural oscillator model using sinusoidal wave and band noise was demonstrated. Through numerical simulations, it was found that band noise stimulus can inhibit the oscillation. Therefore, the finding of this experiment could explain the fact that the human auditory system temporarily halts perception of tinnitus following TM and that it is a kind of sound therapy using band noise stimuli.

It has been reported [27] that for TRT lower amplitudes were used for the stimulus so that the patients could hear the sound of tinnitus. In this current study, the simulations used the Gaussian white-noise stimulus that TRT employs. These simulation results seem to be consistent with the practice [44].

The parameters of plasticity were arbitrarily determined so that the simulation was performed within an appropriate time. The time scale would be much longer in the clinical situation. Further correspondence of the simulation data to clinical data needs to be examined.

Future work will expand this model so that it can more effectively relate to the underlying physiology of tinnitus, and explore better stimulation for its inhibition. This, in turn, will result in improvement in designing better and more effective sound therapy techniques and stimuli. The present model consists of the simplest arrangement of neuronal ensembles that can produce an oscillatory state which would

resemble the perception of tinnitus. As discussed, the arrangement represents the aggregate mechanisms in the thalamo-cortico-thalamic loop and future enhancements of the model will focus on adding excitatory and inhibitory connectivity between units E_1 and I in order to capture intrathalamic interactions.

Acknowledgments

This work was supported in part by Grant-in-Aid for Scientific Research #21560429 from Japan Society of Promotion of Science.

References

- [1] E. P. Fowler, "Head noises in normal and in disordered ears: significance, measurement and treatment," *Archives of Otolaryngology – Head and Neck Surgery*, Vol. 39, No. 6, 1944, pp. 498-503.
- [2] M. J. Penner, "An estimation of the prevalence of tinnitus caused by spontaneous otoacoustic emissions," *Archives of Otolaryngology – Head and Neck Surgery*, Vol. 116, No. 4, 1990, pp. 418-423.
- [3] A. Sismanis and W. R. Smoker, "Pulsatile tinnitus: recent advances in diagnosis," *The Laryngoscope*, Vol. 104, Issue 6, 1994, pp. 681-688.
- [4] A. Axelsson and A. Ringdahl, "Tinnitus – a study of its prevalence and characteristics," *Br J Audiol*, Vol. 23, 1989, pp. 53-62.
- [5] A. J. Heller, "Classification and epidemiology of tinnitus," *Otolaryngol Clin North Am*, Vol. 36, 2003, pp. 239-248.
- [6] J. L. Weissman and B. E. Hirsch, "Imaging of tinnitus: a review," *Radiology*, Vol. 2, 2000, pp. 342-349.
- [7] D. De Ridder, G. De Mulder, V. Walsh, N. Muggleton, S. Sunaert and A. Moller, "Magnetic and electrical stimulation of the auditory cortex for intractable tinnitus, Case report," *J Neurosurg*, Vol. 100, 2004, pp. 560-564.
- [8] W. Muhl nickel, W. Lutzenberger and H. Flor, "Localization of somatosensory evoked potentials in primary somatosensory cortex: a comparison between PCA and MUSIC," *Brain Topogr*, Vol. 11, 1999, pp. 185-191.
- [9] W Meyershoff, "Tinnitus," In: Meyershoff W, Ria D (eds.) *Otolaryngology head and neck surgery*, WB Saunders, Philadelphia, 1992, pp. 435-446.
- [10] J. J. Eggermont, "Central tinnitus," *Auris Nasus Larynx*, Vol. 30 [Suppl], 2003, pp. S7-S12.
- [11] P. J. Jastreboff, "Phantom auditory perception (tinnitus): mechanisms of generation and perception," *Neurosci Res*, Vol. 8, No. 4, 1990, pp. 221-254.
- [12] A. L. Giraud, S. Chery-Croze, G. Fischer, C. Fischer, A. Vighetto, M. C. Gregoire, F. Lavenne and L. Collet, "A selective imaging of tinnitus," *Neuroreport*, Vol. 10, 1999, pp. 1-5.
- [13] W. Muhl nickel, T. Elbert, E. Taub and H. Flor, "Reorganization of auditory cortex in tinnitus," *Proc Natl Acad Sci USA*, Vol. 95, 1998, pp. 10340-10343.
- [14] J. A. Kaltenbach, "Neurophysiologic mechanisms of tinnitus," *J Am. Acad. Audiol.*, Vol. 11, 2000, pp. 125-137.
- [15] D. De Ridder, H. Ryu, A. R. Moller, V. Nowe, P. Van de Heyning, J. Verlooy, "Functional anatomy of the human cochlear nerve and its role in microvascular decompressions for tinnitus," *Neurosurgery*, Vol. 54, 2004, pp. 381-388; discussion pp. 388-390.
- [16] S. A. Chowdhury and N. Suga, "Reorganization of the frequency map of the auditory cortex evoked by cortical electrical stimulation in the big brown bat," *J Neurophysiol*, Vol. 83, 2000, pp. 1856-1863.
- [17] D. Robertson and D. R. Irvine, "Plasticity of frequency organization in auditory cortex of guinea pigs with partial unilateral deafness," *J Comp Neurol*, Vol. 282, 1989, pp. 456-471.
- [18] F. Mirz, B. Pedersen, K. Ishizu, P. Johannsen, T. Ovesen, H. Stodkilde-Jorgensen and A. Gjedde, "Positron emission tomography of cortical centers of tinnitus," *Hear Res*, Vol. 134, 1999, pp. 133-144.
- [19] F. Mirz, A. Gjedde, K. Ishizu and C. B. Pedersen, "Cortical networks subserving the perception of tinnitus – a PET study," *Acta Otolaryngol Suppl*, Vol. 543, 2000, pp. 241-243.
- [20] A. H. Lockwood, R. J. Salvi, M. L. Coad, M. L. Towsley, D. S. Wack and B. W. Murphy, "The functional neuroanatomy of tinnitus: evidence for limbic system links and neural plasticity," *Neurology*, Vol.50, 1998, pp. 114-120.
- [21] W. Arnold, P. Bartenstein, E. Oestreicher, W. Romer and M. Schwaiger, "Focal metabolic activation in the predominant left auditory cortex in patients suffering from tinnitus: a PET study with [18F] deoxyglucose," *ORL J Otorhinolaryngol Relat Spec*, Vol. 58, 1996, pp. 195-199.
- [22] A. Gardner, M. Pagani, H. Jacobsson, G. Lindberg, S. A. Larsson, A. Wagner and T. Hallstrom, "Differences in resting state regional cerebral blood flow assessed with 99mTc-HMPAO SPECT and brain atlas matching between depressed patients with and without tinnitus," *Nucl Med Commun*, Vol. 23, 2002, pp. 429-439.
- [23] A. T. Cacace, "Expanding the biological basis of tinnitus: crossmodal origins and the role of neuroplasticity," *Hear Res*, Vol. 175, 2003, pp. 112-132.
- [24] A. T. Cacace, J. P. Cousins, S. M. Parnes, D. Semnoff, T. Holmes, D. J. McFarland, C. Davenport, K. Stegbauer and T. J. Lovely, "Cutaneous-evoked tinnitus. I. Phenomenology, psychophysics and func-

- tional imaging,” *Audiol Neurotol*, Vol. 4, 1999, pp. 247–257.
- [25] M. Ballester, K. O. Lovblad, A. C. Nirikko, D. Vibert, P. Romanet, G. Schroth and R. Hausler, “Functional MRI of tinnitus – preliminary results using echo-planar imaging (abstract),” *Neuroimage*, Vol. 13, 2001, p. 379.
- [26] J. R. Melcher, I. S. Sigalovsky, J. J. Guinan, R. A. Levine, “Lateralized tinnitus studied with functional magnetic resonance imaging: abnormal inferior colliculus activation,” *J Neurophysiol*, Vol. 83, 2000, pp. 1058–1072.
- [27] M. Muhlau, J. P. Rauschecker, E. Oestreicher, C. Gaser, M. Rottinger, A. M. Wohlshlager, F. Simon, T. Etgen, B. Conrad and D. Sander, “Structural brain changes in tinnitus,” *Cerebral Cortex*, Vol. 16, Sept 2006, pp. 1283-1288.
- [28] J. P. Rauschecker, “Auditory cortical plasticity: a comparison with other sensory systems,” *Trends Neurosci*, Vol. 22, 1999, pp. 74-80.
- [29] N. Suga and X. Ma, “Multiparametric corticofugal modulation and plasticity in the auditory system,” *Nat Rev Neurosci*, Vol. 4, 2003, pp. 783-794.
- [30] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” *Trends in Neurosciences*, Vol. 27, No. 11, 2004, pp. 676-682.
- [31] A. R. Moller, *Neural plasticity and disorders of the nervous system*, Cambridge: Cambridge University Press, 2006.
- [32] T. Tzounopoulos, “Mechanisms of synaptic plasticity in the dorsal cochlear nucleus: plasticity-induced changes that could underlie tinnitus,” *American J. of Audiology*, Vol. 17, Dec. 2008, pp. S170-S175.
- [33] M. Dominguez, S. Becker, I. Bruce and H. Read, “A spiking neuron model of cortical correlates of sensorineural hearing loss: spontaneous firing, synchrony, and tinnitus,” *Neural Computation*, vol. 18, 2006, pp. 2942-2958.
- [34] S. Grossberg, “Linking attention to learning, expectation, and consciousness,” in *Neurobiol. Attention*, L. Itti and J. Tsotsos, Eds., 2005, pp. 652-662.
- [35] D. J. Strauss, W. Delb, R. D’Amelio, Y. F. Low and P. Falkai, “Objective quantification of the tinnitus decompensation by synchronization measures of auditory evoked single sweeps,” *IEEE Trans. Neural Systems and Rehabilitation Eng.*, Vol. 16, Feb. 2008, pp. 74-81.
- [36] C. Trenado, L. Haab, W. Reith and D. J. Strauss, “Biocybernetics of attention in the tinnitus decompensation: an integrative multiscale modeling approach,” *J. Neurosci. Methods*, Vol. 178, 2009, pp. 237-247.
- [37] C. Trenado, L. Haab and D. J. Strauss, “Corticothalamic feedback dynamics for neural correlates of auditory selective attention,” *IEEE Trans. Neural Systems and Rehabilitation Eng.*, Vol. 17, Feb. 2009, pp. 46-52.
- [38] J. A. Henry, M. A. Schechter, T. L. Zaugg, S. Griest, P. J. Jastreboff, J. A. Vernon, C. Kaelin, M. B. Meikle, K. S. Lyons and B. J. Stewart, “Outcomes of clinical trial: tinnitus masking versus tinnitus retraining therapy,” *J. Am. Acad. Audiol.*, Vol. 17, No. 2, 2006, pp. 104-132.
- [39] P. B. Davis, “Music and the acoustic desensitization protocol for tinnitus,” in *Tinnitus Treatment: Clinical protocols*, R. S. Tyler Ed. New York: Thieme, 2006, pp. 146-160.
- [40] R. S. Hallam and L. McKenna, “Tinnitus habituation therapy,” in *Tinnitus Treatment: Clinical protocols*, R. S. Tyler Ed. New York: Thieme, 2006, pp. 65-80.
- [41] K. Fujimoto, H. Nagashino, Y. Kinouchi, A. A. Danesh and A. S. Pandya, “Analysis of a neural oscillator model with plasticity for treatment of tinnitus,” in *Proc. of World Congress on Medical Physics and Biomedical Engineering*, Vol. 14, 2006, pp. 3413-3416.
- [42] K. Fujimoto, H. Nagashino, Y. Kinouchi, A. A. Danesh and A. S. Pandya, “Oscillation and its inhibition in a neural oscillator model for tinnitus,” in *Proc. of the 28th IEEE EMBS Annual International Conference*, 2006, pp. 5547-5550.
- [43] K. Fujimoto, H. Nagashino, Y. Kinouchi, A. A. Danesh and A. S. Pandya, “A plastic neural network model for sound therapy of tinnitus,” *IEEJ Trans. on Electrical and Electronic Engineering*, vol. 2, no. 4, 2007, pp. 488-490.
- [44] K. Fujimoto, H. Nagashino, Y. Kinouchi, A. A. Danesh and A. S. Pandya, “Dynamical properties of a plastic neural network model for tinnitus therapy and inhibition of oscillation using noise stimulus,” in *Proc. of the 29th Annual International Conference of the IEEE EMBS*, 2007, pp. 2408-2411.
- [45] K. Fujimoto, H. Nagashino, Y. Kinouchi, A. A. Danesh and A. S. Pandya, “Inhibition of oscillation in a plastic neural network model for tinnitus therapy using noise stimulus,” in *Proc. of 2007 World Multi-Conference on Systemics, Cybernetics and Informatics*, 2007, pp. 108-112.
- [46] D. O. Hebb, *The Organization of behavior: A neuropsychological theory*. New York: John Wiley & Sons, 1949.

Biographies

HIROFUMI NAGASHINO is currently Professor in Subdivision of Biomedical Information Science, Division of Health Sciences, Institute of Health Biosciences, The University of Tokushima. He received the Bachelor of Engineering and Master of Engineering degrees in Electrical Engi-

neering from The University of Tokushima, Japan in 1972 and 1974, respectively. He received the Doctor of Engineering degree in 1982 from Osaka University, Japan. His research interest includes biocybernetics, neural networks and its application to biomedical engineering, particularly neural network models for oscillatory activities, signal source identification, pattern recognition, etc. Dr. Nagashino may be reached at nagasino@medsci.tokushima-u.ac.jp

KEN'ICHI FUJIMOTO is currently an Assistant Professor in Subdivision of Biomedical Information Science, Division of Health Sciences, Institute of Health Biosciences, The University of Tokushima. He received the Bachelor of Engineering degree in Information Science and Engineering from The University of Tokushima, Japan in 1995 and Master of Engineering degree in Information Science and Engineering from Tokyo Institute of Technology, Japan in 1997. He received the Doctor of Engineering degree from The University of Tokushima, Japan in 2000. His research interests include basic research on the dynamics of nonlinear dynamical systems and its applications to biomedical engineering, image processing, and image reconstruction in computed tomography. Dr. Fujimoto may be reached at fujimoto@medsci.tokushima-u.ac.jp

YOHISUKE KINOUCI is currently a Professor Emeritus and Deputy Director, The University of Tokushima, Tokushima, Japan and is also a Guest Professor at Harbin Institute of Technology, Shenzhen Graduate School, Shenzhen, China. His current research interests include magnetic dentistry, biological effects of magnetic fields, bioimpedance, blood flow measurement, mobile telemedicine, medical applications of neural networks, physiological inverse problems and medical applications of LED. Dr. Kinouchi may be reached at kinouchi@ee.tokushima-u.ac.jp

ALI A. DANESH is currently associate professor in the Department of Communication Sciences and Disorders, Florida Atlantic University. He received Bachelor of Science degree in Audiology from College of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran in 1987 and Master of Science in Audiology from Idaho State University, Pocatello, Idaho, USA in 1994. He completed his PhD in Audiology with an emphasis on Auditory Electrophysiology from the University of Memphis, Memphis, Tennessee, USA in 1998. His research interests include auditory electrophysiology, tinnitus, auditory processing and vestibular disorders. Dr. Danesh may be reached at danesh@fau.edu

ABHIJIT S. PANDYA is currently Professor in the Computer Science and Engineering Department at Florida Atlantic University, Boca Raton, Florida, USA. received his

undergraduate education at the Indian Institute of Technology, Bombay and graduated with a M.Sc. in Physics (specialization in Electronics) in 1977. He earned his M.S. and Ph.D. in Computer Science from the Syracuse University, New York in 1985 and 1988 respectively. His areas of research include VLSI implementable algorithms, Applications of AI and Neural Networks, Image analysis in Medicine and Electronic Health Records. Dr. Pandya may be reached at pandya@fau.edu

HYDROELECTRIC PLANT DESIGN AS A SYNTHESIS FOR AN ENGINEERING TECHNOLOGY CURRICULUM

Henry Foust, Nicholls State University; George Watt, Nicholls State University

Abstract

Hydroelectric plant design is an involved process that includes dealing with hydrology, fluid mechanics of open and pressure conduits, turbomachinery, economics, and electrical power generation. This process involves an economic trade-off condition between capital cost and possible profit from electricity generation; this is an implicit optimization problem. Due to the nature of these types of designs, hydroelectric plant design involves a great deal of engineering judgment. Enhanced ability to make engineering judgments is one desired outcome of any engineering or engineering technology program. This paper deals with a design problem for students in a fluid power course within an engineering technology curriculum and shows the benefits of a final project and that requires some engineering judgment.

Introduction

How do we prepare students for a work environment, when classes are often highly structured and their future work environment is often fraught with ambiguous situations that call for spur-of-the-moment judgments? How do we instill engineering judgment in students? Teachers need to go beyond just giving the students a set of tools to be utilized for a broad set of circumstances by actually showing the students when to use a certain tool, which is the beginning of engineering judgment.

This proposed approach to address these issues was to teach a fluid-power class, during a fourteen-week semester, in the following manner. The class began with ten weeks of fluid mechanics, beginning with a basis in conservation of energy, mass, and momentum. The course then went on to discuss fluid statics, pipe flow, minor losses, pump design, and turbine design. The final four weeks were devoted to the design problem and used a series of four reports on the design of critical elements of a hydroelectric plant. The skills taught by the hydroelectric plant design are:

1. The first skill that is generally conveyed is that the initial design is based on an estimate, a guess as the students would put it.
2. Integration of economics, power concepts, and optimization is required.
3. An optimization is performed to find the trade-off condition between capital cost for the penstock ver-

sus head-loss that affects the present worth analysis.

4. Use of non-dimensional analysis.
5. Using Excel to determine head-loss in penstocks.
6. Determination of specific speed, turbine type and wheel diameter from the Cordier's diagram.
7. Understanding efficiency
8. Engineering judgment

The outline of this paper begins with a background section that explores the nature of manufacturing engineering technology at Nicholls State University, followed by the educational outlook of the authors, and the fluid power class used for this study. The hydroelectric plant design will then be discussed followed by a presentation of a survey and its results. These results, along with the study, are discussed and conclusions made.

Related Research

The manufacturing engineering technology program at Nicholls State University Our manufacturing engineering technology program is built around a core series of project courses that serve as the capstone-manufacturing experience requiring implementation of the technical and business management courses. The novelty of these project courses is that the students from three grade levels (senior, junior, and sophomore) participate in the projects. This approach provides both horizontal and vertical integration of the design/manufacturing project, which is not common [1]. Students register for the appropriate sophomore-, junior- or senior-level project course, but they all meet together and are administered under the same course structure.

The students are divided into several teams with each team functioning as a division in a small company [2-4]; each team participates in a two-semester-long project. The tasks assigned each class level are designed to be compatible with the courses that a typical student should have completed or be taking at the sophomore, junior, or senior levels. Each student gets experience at his or her level of education with the most intense and demanding load being placed on the seniors. The intent of this project approach was to give the students a more realistic engineering experience while they are still at the university so they will be better prepared for their first real-world experiences. Additional approaches that allow the faculty to meet student-learning needs are also being considered. In so doing, the authors have chosen two courses in which to explore some of the relevant teaching

and learning theories were believed to be particularly applicable to engineering technology curricula. The theories are (1) Piaget's intellectual development theory, (2) scientific learning cycle, and (3) Kolb's learning and teaching cycle [5]. A brief discussion of these theories follows.

Educational Outlook

Piaget's theory discusses the stages of intellectual development from new-born to adult. But this study was only concerned with the concrete-operational and formal-operational stages, and the transition from the former to the latter. The concrete-operational stage addresses the ability to do mental operations only with real (concrete) objects, events or situations. The formal-operational stage relates to being able to think abstractly; formulate hypotheses without relating to real objects, events or situations; testing hypotheses and logical alternatives; and generalizing from real objects to abstract notions and ideas. At the formal-operational stage, the student is capable of learning higher math and is capable of applying it to the solution of new problems. Engineering technology education requires at least some ability to operate at the formal-operational level. Students operating at the concrete-operational level learn math by memorization and are usually unable to use it to solve new or unusual problems and, consequently, will have difficulty with an engineering type of curriculum.

Piaget felt that the transition from concrete operational to formal operational was complete by the twelfth year, but more recent studies [6], [7] have shown that as much as 60% of the adult population appears to still be in the transitional stage between the two phases. That is, they can sometimes use formal operational thought, but not always. Experience shows that many of our engineering technology students are in the transition between the concrete- and formal-operational stages of intellectual development. A critical question is "How do we help our students move more toward formal operational intellectual development"? Piaget proposed that the transition is initiated by the introduction of new ideas that don't fit the individual's current mental structures, thus creating a disequilibrium that the individual must deal with in some way. The new ideas may be dealt with by rejecting them or, if they must be dealt with, they are memorized but not understood—this is often what the concrete-operational individual does—if they are not too dissimilar, they may be accommodated into the existing mental structure, if they are quite different from the existing mental structure they may be transformed to fit the mental structure or, ideally, the mental structure is changed and grows to be able to assimilate these new ideas. As the assimilation occurs, the individual's intellectual development stage becomes more formal operational and less concrete operational. Two well-known learning cycles have been proposed that have the potential to create the needed disequilibrium

and help guide students in assimilating the new ideas. They are the Scientific Learning Cycle and the Kolb Learning Cycle. Each will be discussed briefly.

The Scientific Learning Cycle is a three-stage cycle. The first stage is an exploration_or self-discovery phase where students explore a new phenomenon with minimal guidance and try to learn how it works by using it. The exploration phase could involve computer simulations to explore a process or device. Phase two is called the term-introduction or invention phase, or the concept-invention or introduction phase where the instructor fills in the parts the students missed in the exploration phase. And, finally, the concept-application or expansion phase where students apply the new ideas, terms, and/or patterns to new examples by homework, discussion, laboratory exercises, etc. This learning cycle is simple and straightforward in principle, but can take a great deal of preparation by the instructor to have a properly designed self-discovery phase.

The Kolb Learning Cycle is similar, but is based on the idea of two dichotomies considered to be orthogonal to each other. The first dichotomy addresses how individuals transform experience to knowledge and contrasts Active Experimentation (AE) with Reflective Observation (RO). The second is Abstract Conceptualization (AC) versus Concrete Experience (CE) and addresses how individuals grasp or take in knowledge. The two dichotomies are arranged at the extremes on orthogonal axes called Processing Information (horizontal axis) and Taking in Information (vertical axis), as shown in Figure 1.

Complete learning demands that all four stages or steps be covered. And although one may enter the cycle at any point, the most common point to enter is at CE where personal involvement is required. For complex information sets the cycle may be traversed several times in an upward spiral progression. There may be attempts to short circuit the cycle by skipping steps, but this results in incomplete learning. Retention of information as a function of the steps completed is assessed as follows: (1) AC only-20%, (2) AC + RO-50%, (3) AC + RO + CE-70%, and AC + RO + CE + AE-90%. An example of a complete and reasonably good teaching cycle to match the learning cycle would be lecture (RO) followed by requiring that the students think about the ideas presented (exactly how to do this might cause questions) (AE) and then homework (CE) followed by a required laboratory experiment. Lecture (RO) followed only by homework (AE) is not very good. In a continuing emphasis on team work and communication, both written and oral reporting, and most assignments, will require working in teams so that student interactions will aid the learning process and students will develop the ability to communicate and work in teams.

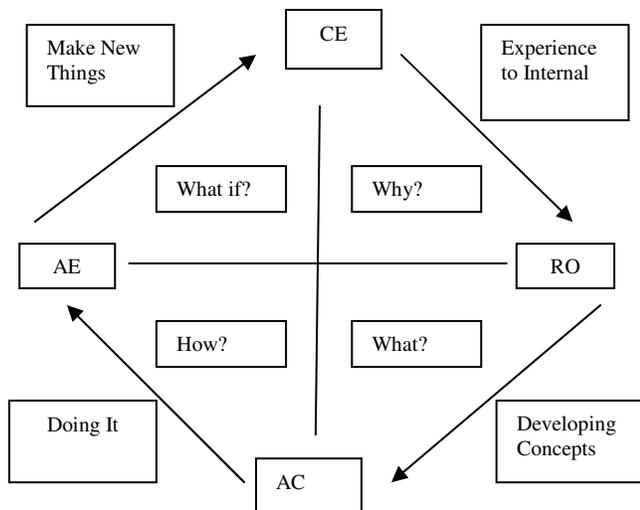


Figure 1. Modified Kolb Learning Cycle

It is also desirable to take the students through all stages of the learning cycle by carefully chosen lectures, often occurring in a just-in-time manner: discussions, discovery experiences, working in teams, analyzing, synthesizing, building, and testing. This emphasis on systems will allow both analyses and syntheses.

The full Kolb Learning Cycle was not employed in this experience, but most of it was followed and, in fact, the scientific learning cycle was approximated very closely. The point is that the first author was involuntarily following some of the principles discussed above. If we use the Kolb Learning Cycle (see Figure 1), he began at the point labeled RO. Through lectures, homework and exams, the students developed abstract conceptualization, AC. After the fact, the laboratory experience brought the students to CE, which occurred throughout the semester. “After the fact” in this case means that motivation should precede all new material, which will be further developed in the discussion.

The final project brought the students toward AE and through to CE. The students did go through the complete cycle because several projects were assigned and completed with minimal guidance. The remainder of the paper describes the details of the approach taken in this attempt to apply the scientific learning cycle method.

Fluid Power and Fluid Power Laboratory

Fluid Power is essentially a fluid mechanics class with an emphasis on hydraulic machines that can do useful work [2], [4], [8], [9]. The class was taught as fluid mechanics with an emphasis on pumps and turbines. Several weeks were devoted to pumps and turbines.

The final four weeks were devoted to designing critical elements of a hydroelectric plant. This project was useful for several reasons. It brings home a plethora of fluid mechanics concepts that all have to be mastered in order to develop a successful design. It integrates engineering economics for a successful design. More importantly, it cultivates engineering judgment.

Problem Statement

The hydroelectric plant design involved four reports. The first was an initial economic design based on a capital cost, annual benefits from electricity generation, and other costs. The second report was the design of a penstock, involving a trade-off condition between capital costs and head-loss. From this second design, five scenarios were determined: 25, 30, 35, 40, and 45cm diameter penstocks. The third report was the design of a turbine and determination of its cost. The fourth report was a final economic analysis with better estimates of the capital costs for the three scenarios.

This section shows the required work delineated into four sections. Figure 2 presents a conceptual model of the hydroelectric plant.

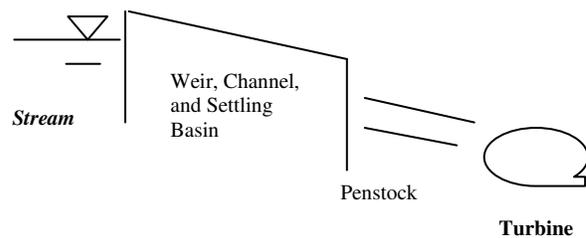


Figure 2. Diagram of Hydroelectric Plant

This was a four-week project to design critical elements of a hydroelectric plant. It should be noted that this design was not complete. Each Friday by close of business a 5-page report would be due. The requirements of each report are given below.

Preliminary Design

1. Determine kWh's produced
2. Determine annual benefits
3. Given useful life of 12 years and the following cost data, is this a viable project at a marginal rate of return (MARR) of 10%? See Table 1.

Table 1. Initial Design Information

Capital Cost	\$285,000
Annual O&M	10% of Capital Cost
Salvage Value	10% of Capital Cost
N	12 years
Electricity Sale Value	\$.06 per k@-hr.
MARR	10%

Penstock Design

1. Given a length of 610 meters, determine an optimal D
2. Determine head-loss for penstock (use 5%, 10%, 15% and 20%) and determine diameters
3. Determine cost of penstock given steel is \$.25 per Newton, which is the cost of the material, emplacement, and supports

Turbine Design

1. Determine N_s given $Q = .30 \text{ m}^3/\text{s}$, Total Head is 91.5 meters, and $N = 600 \text{ rpm}$. Determine $H(a)$ from penstock head-loss and assume 5% head-loss for weir and channel. An approximate value can be that the available head is 85% of the total head.
2. Determine appropriate turbine type from Table 3.2 in handouts (provided to students)
3. Determine D_s from Cordier diagram
4. From D_s , determine D for the wheel
5. Determine the blade angle and nozzle diameter
6. Given head-loss for weir, channel, penstock, and 10% for turbine, and 10% for generator and electrical transmission, determine kW's produced (see Table 2)
7. Determine turbine cost from handout

Table 2. Efficiencies

$\epsilon(\text{weir})$	95%
$\epsilon(\text{channel})$	
$\epsilon(\text{penstock})$	To be determined
$\epsilon(\text{turbine})$	90%
$\epsilon(\text{electrical})$	90%

Final Design

Recalculate the cost analysis for various head-loss cases and turbine cost using Table 3. Note that each case was a different penstock inner diameter and results in different values for Penstock, Turbine, Generator, and Turbine House costs.

Table 3. Cost Information

	Case 1	Case 2	Case 3	Case 4
Civil Works	\$25,000	\$25,000	\$25,000	\$25,000
Penstock	TBD	TBD	TBD	TBD
Turbine	TBD	TBD	TBD	TBD
Generator, Turbine House	Same as Turbine	Same as Turbine	Same as Turbine	Same as Turbine
Misc. Cost	\$20,00	\$20,00	\$20,000	\$20,000

Methodology for the Design

Initial Design

From Table 4, the students discerned that the project may be viable and they proceeded to the next step, which was the penstock design. The students had to estimate capital costs and efficiency (ϵ) in order to calculate the benefits derived from power generation. This leaves the students uncomfortable because essentially this initial design is a guess. This guess is a step in the direction of engineering judgment.

Table 4. Initial Design

Q (cms) =	.30
HT (meters) =	91.5
Capital =	\$285,000
N (years) =	12
Sale (\$/kW-hr)	0.06
O&M	10%
Salvage Value	10%
MARR	10%
Efficiency (ϵ)	50%

Power Generation	
Power (kW) =	135
E/yr (kW-Hr)	1180759
Benefits	70846
Annual O&M	\$28,500
Salvage Value	\$28,500
Present Worth (10%) =	\$21,691

As can be seen in Table 4, the initial design is a viable option at a marginal rate of return of 10% with a present value of \$21,691. It was assumed that the overall efficiency of the plant was 50% and would have a useful life of 12 years.

Penstock Design

The penstock design (Table 5) is a difficult element of the project because there is a trade-off condition between the size of the penstock and the associated capital cost versus the head-loss and power efficiency. Having the students deal with trade-offs and comparing engineering and economic considerations was another important lesson.

As expected, increasing the pipe diameter reduced head-losses, but also increased capital costs. It was found that the optimal diameter to maximize the final economic analysis (see Table 7) was .35 meters.

Table 5. Penstock Design

Diameter (m)	0.25	0.3	0.35	0.4
HL (m)	65.9	26	11.9	6.1
D (Inner) (m)	0.25	0.3	0.35	0.4
D (Outer) (m)	0.26	0.32	0.37	0.42
A (m ²)	0.005	0.007	0.01	0.013
V (m ³)	3.07	4.42	6.02	7.86
Weight (N)	236,914	341,157	464,352	606,501
Cost	59,229	85,289	116,088	151,625

Turbine Design

The turbine design is based on the Cordier Diagram, which is discussed by Logan [4]. The Cordier Diagram is a means of determining the most efficient non-dimensional diameter (Ds) for a specific speed (Ns). Given Ds, the ideal diameter of the turbine can be determined. Based on Ns, the most efficient type of turbine can also be determined where Pelton Wheel is (.03-.3), Francis Turbine is (.3-2.0) and Kaplan Turbine is (2.0-5.0). The analysis is given in Table 6. Figure 3 gives cost data for Pelton Turbines [9]. For brevity, information for Ds and appropriate wheel diameters are not given in Table 6.

Table 6. Turbine Design

H (a) "m"	21.03	60.93	75.03	83.53
N (rpm)	600	600	600	600
N	62.83	62.83	62.83	62.83
N(s)	0.63	0.28	0.24	0.22
Type	Francis	Pelton	Pelton	Pelton
Power (kW)	50	145	179	199
Cost	55,140	79,822	97,811	108,656

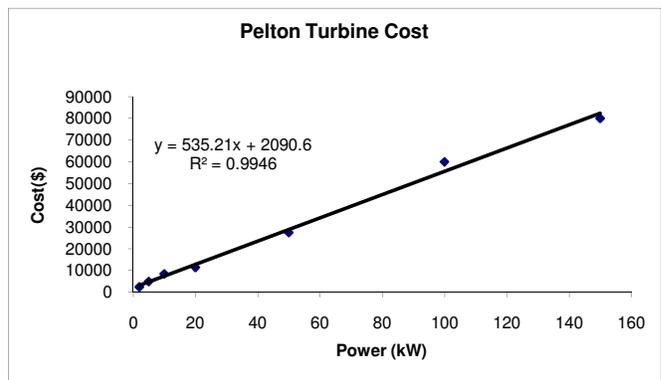


Figure 3. Turbine Cost [5]

Final Design

The final design is given in Table 7 and the present worth of each option is shown in Figure 4. In Figure 4, the options investigated were a penstock with .25 to .50 meter diameters. It is obvious that .35 meters is the best diameter to maximize the present worth.

Table 7. Economics

Case	0.25	0.3	0.35	0.4
Weir	5000	5000	5000	5000
Channel	20000	20000	20000	20000
Penstock	59229	85289	116088	151625
Turbine	55140	79822	97811	105211
Generator	55140	79822	97811	105211
Etc.	20000	20000	20000	20000
Capital Cost	214508	289933	356710	407047
e(overall)	19%	54%	66%	72%
Power (kW)	50	145	179	193
Price	0.06	0.06	0.06	0.06
Benefits	26372	76419	94105	101380
PW	-174142	42450	52806	19344

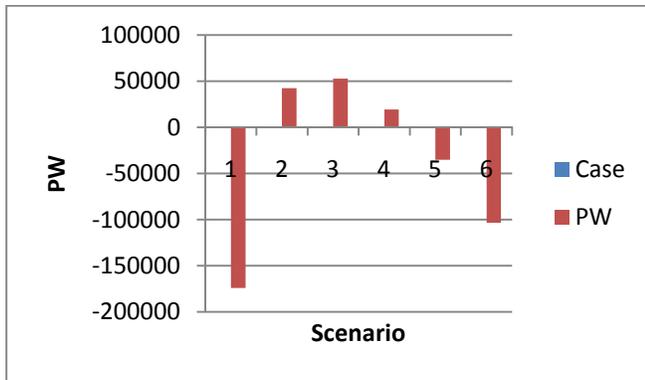


Figure 4. Present Worth Analysis

Discussion

Surveys were given before the last four weeks of class and during the last week of class.

Questions asked included:

- 1) What do you like best about this class?
- 2) What do you like least about this class?
- 3) How well is the teacher doing?
- 4) How well are you doing in this class?
- 5) How well is your team functioning?
- 6) Is the course load just right, too much work, or too little work? Where {1,2,3} is too little, {4,5,6,7} just right, and {8,9,10} too much.
- 7) How well do you feel you know the material?
- 8) What is your overall rating of this class?

The results of the survey for questions 3 through 8 (which have quantitative values) are given below in Table 8. These results will be commented on in the Discussion section.

Table 8. Survey Assessment, where 1 is the worst and 10 is the best

	Q3	Q4	Q5	Q6	Q7	Q8
Before (n = 7)						
Average	6.9	7.1	7.7	6.9	6.4	6.4
Standard Deviation	1.2	1.5	1.9	1.1	1.5	2.0
After (n = 3)						
Average	8.0	8.7	9.0	6.0	8.3	8.0
Standard Deviation	0.0	0.6	0.0	0.0	0.6	0.0

As can be seen from Table 8, the students felt by the end of the last four weeks of the semester that the teacher was doing better, they were doing better, their team was functioning better, the class had gotten slightly easier and that they had a firmer grasp on the material and rated the class highly. It is interesting to note that in both the 'before' and 'after' surveys, the students indicated that they were doing

better in the class than the teacher. This may be particular to the class or it may be particular to student psychology.

The results of this survey have to be questioned, especially given that the 'after' survey only included three students in the population and these were likely better students. These better students would bias the results. Also, are the results of the survey statistically significant?

Other observations from this class include: trade-off conditions are the basis to optimization problems and should be presented to the students early on, because many of their work designs will involve some form of tradeoff; the students did not like the quizzes, which was one of the best assessment tools; the students tended to enjoy the hydroelectric plant design project, which was open-ended, because it allowed them to provide unique solutions.

It needs to be discussed that the hydroelectric design involved some simplifications that should be addressed. The assumed efficiencies for the weir, channel, settling basin, turbine and electrical generation need to be explored in an actual design; the electrical components of the system were ignored, but are very important; cost data was from 1990 and an actual design would involve vendor quotes for equipment and services; and, the sales cost of electricity generated would have to be assessed for the area where the electricity was generated.

Further Research

The fluid power course utilized for this work was a course with a problem-based learning pedagogy. The hydroelectric plant design utilized in this paper is one possible final project. Other final projects could be the design of a pump/pipe system that includes a cost estimate, other aspects of a hydroelectric plant not covered in this work, or a compressor/pipe system design with detail cost estimations.

Future educational research papers could explore how an instructor presents an open-ended final project and guides the students through the project, while having the student's take ownership of the work. One idea on how this could be done is to provide the students with an RFP of work, have the student groups propose the work, have the instructor act as the party asking for bids and limited to the knowledge that party would have, and have a third-party act as a senior engineer to guide the students through the project.

Conclusion

There is some benefit to utilizing an end-of-the-semester, open-ended problem that incorporates many seemingly disparate ideas and utilizes a systems approach. The benefits

include: allowing the students to discover engineering concepts, which empowers them; synthesis of many ideas; for this particular problem, incorporation of economics and explicitly addressing the trade-off of capital cost versus yearly revenues from power generation.

Better assessment tools need to be utilized. A more thorough method of assessing teaching and learning methods is being conducted in the upcoming semester. Assessment includes several surveys throughout the semester (both written and oral). The oral surveys will be conducted by a student who has taken the class. Utilizing graded homework and tests to assess different approaches to learning and teaching methods, and a final project to bring it all home.

References

- [1] Alabart, J.R., Herrero, J., Giralt, F., Grau, F.X., and Medir, M., "Two Way Integration of Engineering Education Through a Design Project," *Journal of Engineering Education*, April 2000, pp. 219-229.
- [2] Foust, Henry (2005). Class notes for EGTC 245 and EGTC 246; Nicholls State University.
- [3] Humphreys, M.A., Meier, R.L., and Williams, M.R., "Refocusing Our Efforts: Assessing Non-Technical Competency Gaps," *Journal of Engineering Education*, July 2000, pp. 377-385.
- [4] Logan, Earl (1993). Turbomachinery: Basic Theory and Applications. Marcel Dekker, New York, NY.
- [5] Pavelich, M.J. (1984). "Integrating Piaget's principles of intellectual growth into the engineering classroom." *Proceedings ASEE Annual Conference*, ASEE Washington, DC, 719 to 722.
- [6] Pintrich, P.R. (1990). "Implications of psychological research on student learning and college teaching teacher education" in W.R. Houston, M. Haberman, and J. Sikula (Eds.), *Handbook of Research on Teacher Education*, MacMillan, New York, 926 to 857.
- [7] Granger, Robert (1995). Fluid Mechanics. Dover Publishing; New York, NY.
- [8] Harvey, Adam (1993). Microhydro Design Manual; Intermediate Technology Publications, Bourton-on-Dunsmore, Rugby, UK.
- [9] Roberson, J. and Crowe, C. (1997). Engineering Fluid Mechanics; John Wiley and Sons; New York, NY.

Biographies

HENRY FOUST is an assistant professor in the Department of Applied Sciences at Nicholls State University. His terminal degree (Ph.D., Tulane University) is in Civil Engineering, but he's currently working within a Manufacturing Engineering Technology program where he teaches engi-

neering science courses. His research interests include multiphase flow phenomena and nonlinear control. Dr. Foust can be reached at henry.foust@nicholls.edu.

GEORGE WATT is an associate professor in the Department of Applied Sciences at Nicholls State University. His terminal degree (Ph.D., Ohio State University) is in Metallurgical Engineering and he teaches manufacturing courses within a Manufacturing Engineering Technology program. His research interests include engineering education, material failure and corrosion. Dr. Watt can be reached at george.watt@tractiontechnologiesllc.com.

COMPUTATION OF SHOCKWAVE STRUCTURES IN WEAKLY IONIZED GASES BY SOLVING BURNETT AND MODIFIED RANKINE-HUGONIOT EQUATIONS

Xiaoqing Qian, Z.T. Deng Mechanical Engineering Department Alabama A&M University

Abstract

The modified Rankine-Hugoniot equations across a standing normal shockwave were discussed and adapted to obtain jump conditions for shockwave structure calculations. Coupling the modified Rankine-Hugoniot equations with the Burnett equations, the shockwave structure in a weakly ionized gas flow was computed and analyzed for a wide range of free-stream Mach numbers ranging from 1.75 to 6.0, with ionization ratio ranges from 0 to 5 parts per million. Results indicated that the modified Rankine-Hugoniot equations for shockwave structures in weakly ionized gas are valid for a small range of ionization fractions at low free-stream Mach numbers. The jump conditions also depend on the value of free-stream pressure, temperature and density. The computed shockwave structure with ionization indicated that by the introduction of the weakly ionized gas particles in the main flow field, shockwave thickness was slightly increased and shockwave strength could be reduced.

Introduction

Over the past two decades, scientists [1], [2] have observed that hypersonic shockwaves could be altered by weakly ionizing the air flows. Shadowgraph experiments conducted at AEDC [3] showed that with a small amount of pre-ionization—about one part per million background ionization fraction—when shockwaves pass through the weakly ionized air, a normal Mach 6 shockwave was transformed into a Mach-3-like shockwave. The shockwave stand-off distance was also increased. Other researches [4], [5] indicated that with small ionization levels, typically at about one part in a million by mass, the strength of the bow shock ahead of the supersonic projectile such as sphere was reduced, the net drag on the sphere was reduced, and the shockwave stand-off distance increased. In addition to the effect of gross shockwave properties, the shock structure was also modified including a weakening shock front and a broadening shockwave thickness [5]. The remaining question was how the weakly ionized gas could achieve drag reduction and how it could alter the shockwave structure.

Shockwaves are regions of gas flow with discontinuities or strong gradients in pressure, temperature, density and velocity. The strength of the shockwave can be evaluated by the pressure ratio, density ratio and temperature ratio across

the normal shockwave. The analytical and experimental study of shockwave structures in a neutral gas provides vital information on fluid physics. Shockwave structure determines shock strength, shockwave thickness, shock standoff distance and heat transfer rate. Figure 1 shows the schematic drawing of detailed shockwave structure in a neutral gas.

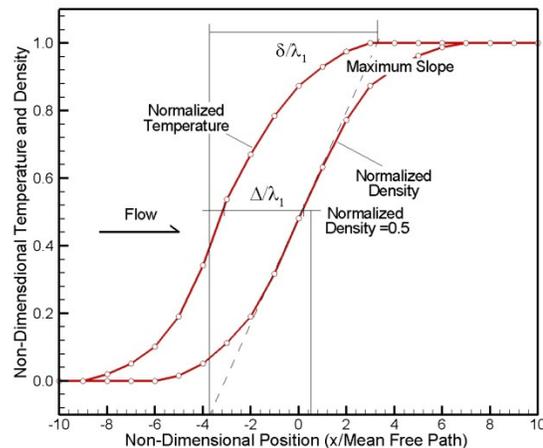


Figure 1. Schematic drawing of shockwave structure in neutral gas

An ideal shockwave has a thickness of zero. However, in reality, with viscous effects, shockwave thickness is usually a few mean-free paths thick. Taking a tangent to the normalized density curve through the shockwave, at the location with maximum slope, the non-dimensional shockwave thickness (δ/λ_1) can be determined as in Figure 1, where λ_1 is the molecular mean-free path of the free-stream gas. The normalized density, ρ_n , and normalized temperature, T_n , inside the shockwave are defined as

$$T_n = \frac{T(x) - T_1}{T_2 - T_1}, \quad \rho_n = \frac{\rho(x) - \rho_1}{\rho_2 - \rho_1} \quad (1)$$

In equation (1), the sub-index 1 denotes conditions ahead of shockwave and 2 denotes those conditions downstream of the shockwave. The separation of the non-dimensional density and temperature profile is called the temperature-density shift, Δ , measured at 50% of the non-dimensional density profile location. Shockwave thickness is given as δ .

Mathematically, the Boltzmann equation is the true governing equation for kinetic fluid dynamics for all flow regimes. Unfortunately, due to the complexity of solving the

Boltzmann equation directly, the continuum Euler and Navier-Stokes equations were used to solve the fluid dynamics problems. The Euler, Navier-Stokes and Burnett equations were derived from the Hilbert-Chapman Enskog expansion of the Boltzmann equation [6-10]. Based on this expansion, the velocity distribution function f was expressed in the series expansion of Knudsen number as

$$f = f^{(0)} + f^{(1)} + f^{(2)} + f^{(3)} + \dots,$$

$$f = f^{(0)}(1 + a_1 K_n + a_2 K_n^2 + a_3 K_n^3 + \dots)$$

which is the perturbation expansion of the velocity distribution function about the Maxwellian distribution, where $f^{(0)}$ is the Maxwellian distribution function, a_i ($i=1,2,3,\dots$) are functions of density, molecular velocity, and temperature. The Knudsen number, K_n , is defined as the ratio of the mean free path of a particle (λ) divided by the characteristic distance (L) over which the macroscopic variables change appreciably, $K_n = \lambda/L$. In this expansion, the Knudsen number, K_n , a small perturbation parameter, must be less than 1.0. Flow regimes can be defined by the Knudsen number. If K_n is much smaller than one, then the Navier-Stokes equations are valid. If K_n is in the range of 0.01-1.0, the flow is typically in the transitional regime [11-13]. In general, the convergence of this expansion is asymptotic as the Knudsen number goes to zero. Substituting this expansion into the Boltzmann equation, taking moments of the Boltzmann equation yields the continuum equations of fluid mechanics, a set of conservation equations describing global conservation of density, momentum, and energy. To close this system of equations requires constitutive equations, which express viscous stress and heat flux in terms of the distribution function rather than macroscopic gradients. The conventional Euler equations are the zeroth-order approximation, the Navier-Stokes equations are the first-order approximation and the Burnett equations are the second-order approximation. As the Knudsen number increases, the Navier-Stokes equations gradually deteriorate because the transitional non-equilibrium effect prevails [11], [12]. It is natural to consider the Boltzmann equation as the governing equation for the transitional flow problems. However, the full Boltzmann equation is very difficult to solve because the collision term is very complex physically as well as numerically.

The flow across the shockwave is in a highly transitional regime where the Knudsen number is relatively large. Typically, shockwave thickness is on the order of a few mean-free path. If the flow characteristic length, L , is 20 times the mean-free path, λ , then the Knudsen number will be in the range of 0.05. It is necessary to consider the high Knudsen-number effects when solving flow characteristics through shockwave. It was shown that Burnett equations are higher-order in accuracy in the transitional regime compared to the

conventional Navier-Stokes equations, and they provide better accuracy in shock-structure prediction [11-13]. The numerical procedure [12], [13] to solve the Burnett equations is explicit. The Burnett terms and boundary conditions were calculated using the Navier-Stokes solution as an initial value. The Burnett stress and heat-flux terms were then treated as source terms and added to the Navier-Stokes equations in order to finally solve them and obtain the Burnett solution. The Burnett solution is a perturbation solution of the Navier-Stokes equations.

In the shockwave-structure prediction analysis, the classical Rankine-Hugoniot equations [14] were applied to describe discontinuous conditions across the shockwave. The classical Rankine-Hugoniot equations were derived from the laws of conservation of mass, momentum and energy for a control volume involving a standing normal shockwave. It relates the densities and pressures for the perfect gas in front of and in back of the normal shockwave.

In a weakly ionized gas flow, pre-ionized charged particles existed in addition to the natural particles present in the natural gas. The high-temperature electrons may excite the vibration modes of molecules, leading to a vibration temperature much higher than the neutral gas temperature [15]. These charged particles can affect the neutral gas-flow characteristics. Shockwaves in weakly ionized gas should exhibit different features from shocks in neutral gases. When a shockwave is present in the weakly ionized gas, large gradients exist in both charged particles (ions and electrons) and neutral gas particles for density. Since electrons move much faster than ions or heavy neutral particles, they will diffuse much faster than ions or neutral particles. Ions will remain near the shock, but electrons will diffuse away from the shock. Also, ions and electrons will be separated and form an electrical double layer across the shockwave [16]. This layer produces an electrostatic body force at the shock directed against the flow as schematically indicated in Figure 2. As a result, the thickness of the shockwave in weakly ionized gas is much wider than the shockwave thickness in neutral gas. The pressure jump across the shockwave in weakly ionized gas is reduced when compared to the jump in neutral gas [16].

To understand the effects associated with the interaction of electromagnetic forces and electrically conducting fluid flow requires integration of several disciplines such as fluid dynamics, electrostatics, chemical kinetics, and atomic physics. Extensive experimental data and simulations are essential for resolving controversial issues. Research attention has been focused on the numerical simulation model development [15-17]. The development of an accurate and robust numerical scheme for weakly ionized gas-flow problems faces challenges in the coupling of the Maxwell equations

with the Navier-Stokes equations through the interaction of electromagnetic fields with the momentum and energy equations.

Shockwave-structure research in weakly ionized flow has mainly been conducted by solving the coupling of the plasma and fluid dynamics Euler equations. The Euler equations were solved for shock structure [16]. A simple model for shockwave jump conditions was proposed and the classical Rankine-Hugoniot relations across the shockwave were modified to include electrostatic force produced by ions and electrons. The effect of weakly ionized gas on shockwaves was analyzed using an electrostatic body force derived from Coulomb's law across the shockwave. The shockwave thickness was approximated by a hyperbolic tangent function. It was assumed that shockwave thickness was about one mean-free path if the free-stream Mach number was above 5. However, in reality, shockwave thickness is unknown prior to the calculation, and Euler equations cannot provide an accurate shockwave structure even without the presence of weakly ionized gas. Shockwave structures cannot be accurately solved using the technique proposed by Seak and Mankowski [16] if shockwave thickness was not given.

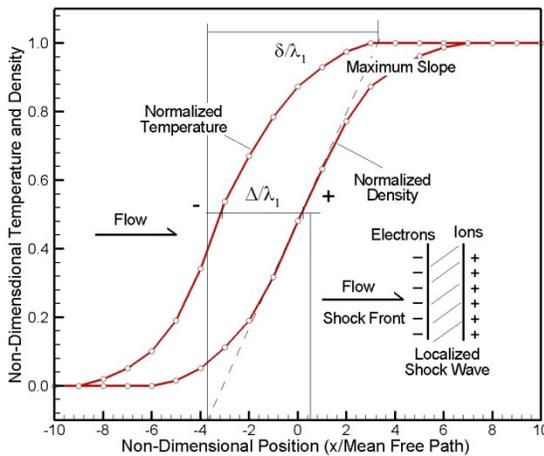


Figure 2. Schematic shockwave structure in weakly ionized gas. Electrostatic body force exists

In this study, the shockwave structure in the weakly ionized gas flow was studied. The objective of this study was to develop a way to numerically investigate the shockwave structure in a weakly ionized gas flow by solving Burnett equations coupled with modified Rankine-Hugoniot relations.

Mathematical Model Description

The shockwave structure in weakly ionized gas is computed by solving the Burnett equations with the addition of

weak ionization effects. It was assumed that the interaction of the weakly ionized gas with fluid could be represented by an electrostatic force acting against the flow direction. It was also assumed that the electrostatic body force could be modified to predict the jump conditions across the shockwave for shockwaves in a weakly ionized gas [16]. Combining the modified Rankine-Hugoniot relations with the Burnett equations, the shockwave structure can be solved in three steps: 1) apply modified Rankine-Hugoniot relations to compute jump conditions (pressure, temperature, velocity and density) across a normal shockwave, where the jump conditions depend on the ionization fraction; 2) determine the magnitude of the electrostatic force acting on the shockwave region against the flow direction, based on the ionization ratio and jump conditions. Since the coupling of the electrostatic field with shock discontinuity is relatively weak for a low-ionization fraction, it was assumed that an averaged electrostatic body force within 80 mean free paths in the direction against the flow inside the shockwave could be used. And, 3) modify the Burnett equations with an electrostatic body force as a source term in order to predict shockwave structures in a weakly ionized gas.

One-Dimensional Burnett Equations:

In conservation form, the one-dimensional Burnett equations with body force per unit mass, f , in Cartesian coordinates can be written as [13], [18]

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} = 0 \quad (2)$$

$$\frac{\partial(\rho u)}{\partial t} + \frac{\partial}{\partial x}(\rho u^2 + p - \sigma) = \rho f \quad (3)$$

$$\frac{\partial(\rho e_t)}{\partial t} + \frac{\partial}{\partial x}(\rho u e_t + pu - \sigma u + q) = \rho f u \quad (4)$$

where,

$$p = (\gamma - 1)\rho(e_t - \frac{u^2}{2}) \quad (5)$$

and where ρ is density, u is the velocity component in the x direction, p is the static pressure, σ is the viscous stress component, q is the heat flux component, and e_t is the specific total energy. The Burnett stress tensor and the heat flux vector component are

$$\sigma = -\frac{4}{3}\mu \frac{\partial u}{\partial x} + \sigma_{Burnett} \quad (6)$$

$$q = -k \frac{\partial T}{\partial x} + q_{Burnett} \quad (7)$$

where

$$\begin{aligned} \sigma_{Burnett} = & \frac{\mu^2}{p} \left(\alpha_1 \left(\frac{\partial u}{\partial x} \right)^2 + \alpha_2 R \frac{\partial^2 T}{\partial x^2} + \alpha_3 \frac{RT}{\rho} \frac{\partial^2 \rho}{\partial x^2} \right. \\ & \left. + \alpha_4 \frac{RT}{\rho^2} \left(\frac{\partial \rho}{\partial x} \right)^2 + \alpha_5 \frac{R}{\rho} \left(\frac{\partial \rho}{\partial x} \frac{\partial T}{\partial x} \right) + \alpha_6 \frac{R}{T} \left(\frac{\partial T}{\partial x} \right)^2 \right) \quad (8) \end{aligned}$$

$$q_{Burnett} = \frac{\mu^2}{\rho} \left(\gamma_1 \frac{\partial u}{\partial x} \frac{\partial T}{\partial x} + \gamma_2 \frac{\partial^2 u}{\partial x^2} + \frac{\gamma_3}{\rho} \frac{\partial u}{\partial x} \frac{\partial \rho}{\partial x} \right) \quad (9)$$

and where k is the thermal conductivity, R is the gas constant, T is temperature, and μ is kinetic viscosity of the gas. The expressions for coefficients for α and γ in equations (8) and (9) can be expressed as [13]

$$\begin{aligned} \alpha_1 &= \frac{2}{3}\omega_1 - \frac{14}{9}\omega_2 + \frac{8}{27}\omega_6 \\ \alpha_2 &= -\frac{2}{3}\omega_2 + \frac{2}{3}\omega_3, \quad \alpha_3 = -\frac{2}{3}\omega_2, \quad \alpha_4 = \frac{2}{3}\omega_2 \\ \alpha_5 &= -\frac{2}{3}\omega_2 + \frac{2}{3}\omega_4, \quad \alpha_6 = \frac{2}{3}\omega_5 + \frac{2}{3}\omega_4 \\ \gamma_1 &= \theta_1 + \frac{8}{3}\theta_2 + \frac{2}{3}\theta_5 \\ \gamma_2 &= \frac{2}{3}\theta_2 + \frac{2}{3}\theta_4, \quad \gamma_3 = \frac{2}{3}\theta_3 \end{aligned} \quad (10)$$

In this study, Maxwellian gases were assumed to obtain these coefficients. The values of ω and θ in equation (10) can be obtained [13] as shown here:

$\omega_1 = \frac{10}{3}$	$\theta_1 = \frac{75}{8}$
$\omega_2 = 2$	$\theta_2 = -\frac{45}{8}$
$\omega_3 = 3$	$\theta_3 = -3$
$\omega_4 = 0$	$\theta_4 = 3$
$\omega_5 = 3$	$\theta_5 = \frac{117}{4}$
$\omega_6 = 8$	

Modified Rankine-Hugoniot Equations:

Across the normal shockwave, the modified Rankine-Hugoniot equations for a normal shockwave can be written as [16]

Continuity:

$$\rho_1 u_1 = \rho_2 u_2 \quad (11)$$

Momentum:

$$p_1 + \rho_1 u_1^2 = p_2 + \rho_2 u_2^2 - \rho_2 f \delta_s \quad (12)$$

Energy:

$$C_{p1} T_1 + \frac{1}{2} u_1^2 = C_{p2} T_2 + \frac{1}{2} u_2^2 - f \delta_s \quad (13)$$

where f is the averaged electrostatic body force per unit mass on shockwave, and δ_s is the shockwave thickness, or the distance required for 99% of jump across the shockwave. The negative body force was used because the coulomb

force on the shockwave due to the ion and electron-space charge in and behind the shock was in a direction opposite to the flow. The sub-indexes 1 and 2 denote conditions before and after the shockwave. The electrostatic body force term can be approximated by multiplying the Coulomb force by ion density at the inner surface of the shock front [16],

$$\rho_2 f \delta_s = k e^2 G_{sc} \left(\frac{\varphi \rho_2}{m} \right)^{\frac{5}{3}} \quad (14)$$

where k is the Boltzmann constant, e is the charge of an electron, G_{sc} is a constant incorporating the size and shape of the space charge region, the non-uniformity of the space charge distribution and the inverse of distance square effect, φ is the ionization fraction, which is the ratio of ion number density to the neutral gas number density, and m is the mass of the atom. To determine the electrostatic force, it was necessary to know the shockwave thickness, δ_s . However, the shockwave thickness is unknown prior to the shockwave structure solution. The selection of δ_s becomes critical for the solution accuracy of the electrostatic force. In neutral gases, shockwave thickness is usually in the range of a few molecular mean free paths [13]. In weakly ionized gas, however, the distance between ions and electrons may be many times the shockwave thickness. A complicated full plasma-dynamics equation has to be solved iteratively to compute the electrostatic force. It was shown [16] that under non-equilibrium conditions, the distance between ions and electrons can be 80 or higher of the mean-free paths. Implicit in this study, as the first step in solving shockwave structures in weakly ionized gas using this modified Rankine-Hugoniot model, the distance between the ions and electron layers across the shockwave was assumed to be 80 mean free paths of the free-stream gas, and the averaged electrostatic body force f in equation (14) can be written as

$$f = \frac{k e^2 G_{sc}}{80 \lambda_1} \left(\frac{\varphi \rho_2}{m} \right)^{\frac{5}{3}} \quad (15)$$

where λ_1 is the mean free path before the shockwave, which can be determined as

$$\lambda_1 = \frac{16 \mu_1}{5 \rho_1 \sqrt{2 \pi R T_1}} \quad (16)$$

In equation (16), μ_1 is the free-stream gas viscosity and R is the gas constant.

Numerical Procedures

The solution technique for the Burnett equations was accomplished through the explicit four-stage Runge-Kutta time integration in time [13]. The convection terms in governing equations (2-4) are discretized using second-order upwind

flux differencing with Roe's nonlinear flux limiter. The first-order Navier-Stokes viscous terms, the second-order Burnett viscous terms and the electrostatic body forces are treated as source terms to the discretized equations and are discretized by using a central differencing technique. The jump conditions across shockwaves in argon gas with weak ionization, such as pressure ratio, temperature ratio and density ratio were obtained by solving the modified Rankine-Hugoniot equations (11-13). This set of non-linear equations was solved using the Newton Raphson's iteration technique. The solution of the modified Rankine-Hugoniot equations provides boundary conditions to the Burnett equations.

Results and Discussion

Modified Rankine-Hugoniot Equations:

The jump conditions were obtained across normal shockwaves by solving the Modified Rankine-Hugoniot equations. Note that the jump conditions depend on the free-stream pressure, temperature and density. In this study, free-stream pressure, temperature and density were selected based on the standard air table at 10,000meter altitudes. Argon gas was used for the calculation. Table 1 shows the computed pressure ratio, temperature ratio and Mach number after the shockwave in weakly ionized argon gas with a free-stream Mach number of 6. In these results, weakly ionized effects were introduced through the volume-weighted ionization fraction,

$$\beta = G_{sc}^{\frac{3}{5}} \phi \quad (17)$$

Notice that when $\beta = 0$, the gas is neutral and no ionization effects exist. Results are the same as standard normal shock table values.

Table 2 shows the computed results for a shockwave Mach number of 3. In both cases, as β increases, the pressure ratio and temperature ratio decrease for a given free-stream Mach number, M_f . As indicated in Table 1, as β increases to 5.0, the Mach number after the shockwave increases to 0.983, which is close to 1.0 for a free-stream Mach number of 3. However, as seen in Table 2, the Mach number after the shockwave was higher than 1.0 when β was larger than 2.25E-06, which was clearly physically impossible.

Normal shockwave theory requires that an incoming free-stream Mach number has to be bigger than 1.0 and a Mach number after a shockwave should be less than 1, in order not to violate the second law of thermodynamics. These results indicated that the simplified Modified Rankine-Hugoniot equations would only be valid for a small value of ionization fraction for moderate free-stream Mach numbers (around 3.0-5.0).

Table 1. Computed jump conditions in weakly ionized argon gas (free-stream Mach number $M_f=6$)

$\beta \times 10^6$	P_2/P_1	T_2/T_1	M_2
0.00	44.750	12.120	0.467
0.25	44.401	12.082	0.470
0.50	43.657	12.000	0.476
0.75	42.641	11.885	0.485
1.00	41.422	11.744	0.496
1.25	40.048	11.580	0.510
1.50	38.560	11.395	0.525
1.75	36.993	11.193	0.543
2.00	35.372	10.975	0.562
2.25	33.720	10.742	0.583
2.75	30.393	10.238	0.632
3.00	28.743	9.968	0.659
3.25	27.113	9.689	0.689
3.50	25.511	9.399	0.721
3.75	23.940	9.099	0.756
4.00	22.404	8.790	0.794
4.25	20.904	8.471	0.835
4.50	19.440	8.142	0.881
4.75	18.011	7.803	0.931
5.00	16.616	7.452	0.986

Table 3 shows the pressure, temperature and Mach number after shockwave for free-stream Mach numbers between 1.75 and 6.0, for ionization fractions of 0.75 parts per million. Figure 3 shows the complete set of computed pressure and temperature ratios across shockwave-versus-ionization fractions at free-stream Mach numbers between 1.75 and 6.0. Notice that when the free-stream Mach number was below 1.75, the modified Rankine-Hugoniot equations failed to accurately predict jump conditions across the shockwave if the ionization ratio was larger than 0.75E-06. This result shows another accuracy problem associated with the modified- Rankine-Hugoniot equations.

Table 2. Computed jump conditions shockwave in weakly ionized argon gas (shockwave free-stream Mach number is 3)

$\beta \times 10^6$	P_2/P_1	T_2/T_1	M_2
0.00	11.000	3.667	0.522
0.25	10.737	3.631	0.532
0.50	10.197	3.555	0.555
0.75	9.499	3.451	0.587
1.00	8.712	3.324	0.628
1.25	7.881	3.178	0.679
1.50	7.039	3.015	0.740
1.75	6.200	2.835	0.815
2.00	5.373	2.637	0.907
2.25	4.552	2.415	1.024
2.50	3.708	2.156	1.188

Table 3. Computed jump conditions shockwave in weakly ionized argon gas at different Mach numbers, $\beta=0.75E-06$

M_1	P_2/P_1	T_2/T_1	M_2
6.0	42.641	11.885	0.485
5.0	29.012	8.449	0.501
4.5	23.158	6.966	0.513
4.0	17.950	5.639	0.529
3.5	13.395	4.467	0.552
3.0	9.499	3.451	0.587
2.5	6.266	2.588	0.642
2.0	3.682	1.868	0.742
1.75	2.607	1.550	0.836

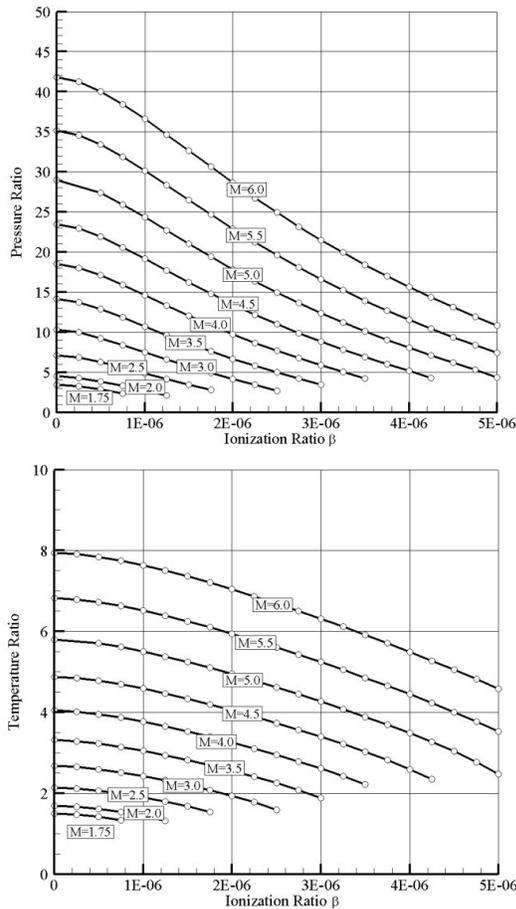


Figure 3. Computed pressure and temperature ratios across normal shock for free-stream Mach numbers ($M=1.75-6.0$) and ionization ratio β

Shockwave structure computation:

The normal shockwave was initially placed in the middle of the computational domain. Over time, the shockwave structure changed. The steady state solution was captured when the four-step Range-Kutta time integration converged. Upstream flow conditions were fixed during the calculation. The downstream pressure condition was fixed and was de-

termined by the solution of the modified Rankine-Hugoniot equations for a given ionization fraction. Other downstream parameters were updated at the downstream exit. The computational domain was 100 times the free-stream molecular mean free path. An averaged electrostatic body force was acting upon this domain for simplicity, as indicated in equation (15). All computations were performed for argon gas. The calculated free-stream Mach numbers ranged from 1.75 to 6.0. Upstream pressure and temperature were chosen from standard air tables at 10,000meter altitudes. Figure 4 shows the normalized temperature and density profile extracted from the shockwave structure solutions using equation (1) inside the shockwave. The free-stream Mach number for this calculation was 5, and the ionization factor, β , was 0.75 parts per million. Physically, the temperature-density shift area was related to entropy change across the normal shock. The horizontal axis was plotted using a non-dimensional scale with reference to free-stream mean-free path.

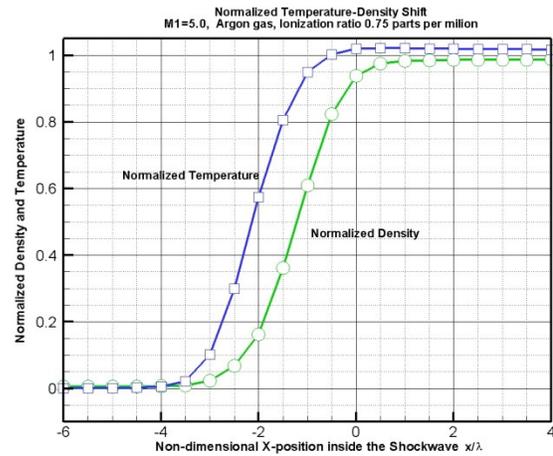


Figure 4. Non-dimensional temperature-density profile for $M_1=5.0$, argon gas, ionization fraction is 0.75 parts per million

Figure 5 compares the temperature-density shift between ionized argon and neutral argon gas. The "with ionization" curve represents the solution with ionization of 0.75 parts per million, while the "without ionization" curve represents the neutral argon gas. A small difference was found at the mid-point of the normalized density profile, where $\rho_n = 0.5$. This indicated that weak ionization does not significantly change the increase in entropy across the shockwave. The entropy change is primarily caused by viscous effects. Figure 6 shows the comparison of the reciprocal shockwave thickness for free-stream Mach numbers between 1.75 and 6.0, between ionized and neutral argon. The solutions were obtained by solving the Burnett equations with and without weak ionization effects. "BUR" represents solution of the Burnett equations without ionization; "ION" represents the solution of Burnett equations with ionization effects. The ionization fraction, β , was 0.75 parts per million for all cas-

es. The reciprocal shockwave thickness (λ_1/δ_s) was measured from the computed, normalized density profile, as in Figure 4, for Mach 5 and for each Mach number using the normalized density profile in equation (1) and as described in Figure 4.

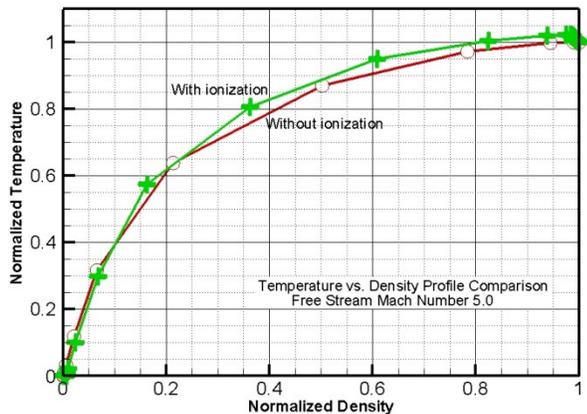


Figure 5. Comparison of the normalized temperature-density shift between weakly ionized gas and neutral gas. Free-stream Mach number is 5, ionization fraction is 0.75 parts per million

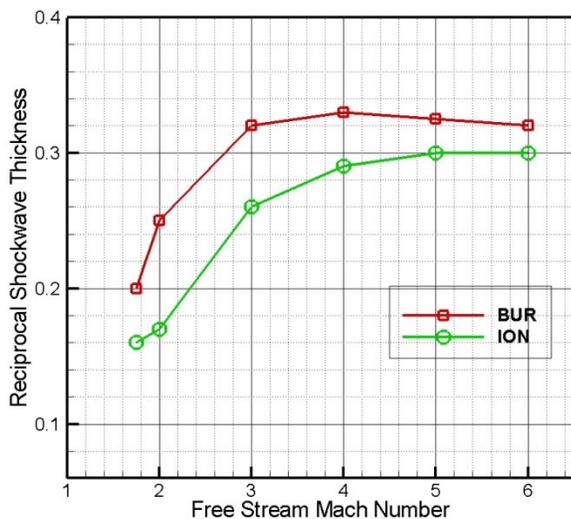


Figure 6. Comparison of reciprocal shockwave thickness (λ_1/δ_s) between ionized argon and neutral argon gas. The ionization fraction is 0.75 parts per million

As indicated in Figure 6, λ_1/δ_s decreases with the existence of weakly ionized gas and, in turn, the shockwave thickness, δ_s , increases with the existence of weakly ionized gas. This demonstrated that the solution of Burnett equation coupling with the modified Rankine-Hugoniot equations can provide a quick and simplified estimate for the reduction of the shockwave strength due to weak gas ionization. This weak ionization shockwave strength reduction was achieved when the ionization ratio, β , was the typical 0.75 parts per million. For low free-stream Mach numbers with high ioni-

zation effects, as indicated in Figure 3, the solution of the modified Rankine-Hugoniot equations (11-14) fail to provide valid jump conditions across shockwaves. Extensive numerical analysis and experimental validation is needed in order to fully understand the validity and the limitation of the modified Rankine-Hugoniot equations.

Results and Discussion

The modified Rankine-Hugoniot equations across a standing normal shockwave were discussed and adapted to obtain jump conditions for shockwave structure calculations. Coupling an electrostatic body force to the Burnett equations, the weakly ionized shockwave structure was solved for a wide range of free-stream Mach numbers between 1.75 and 6.0 with modest ionization ratios. Results indicated that the modified Rankine-Hugoniot equations for shockwaves were valid for a small range of ionization fractions. This model failed to accurately predict valid jump conditions for free-stream Mach numbers below 1.75 and ionization ratios above 0.75E-06. As free-stream Mach numbers increase, the modified Rankine-Hugoniot equations can provide valid jump conditions across a normal shockwave for gas with slightly higher ionization ratios. The computed results of shockwave structure with ionization indicated that shockwave strength may be reduced by the existence of weakly ionized gas. However, further numerical simulation and experimental validation using the modified Rankine-Hugoniot equations is needed. Further studies are also needed to analyze the effects of weakly ionized flow electrostatic force on the change of the specific heat ratio.

References

- [1] Gennadiy I Mishin, Sonic and ShockWaves Propagation in Weakly Ionized Plasma, *Gas Dynamics*, Commack, New York: Nova Science Publishers, Inc. 1992.
- [2] Klimov, A. I., Koblov, A.N., and Mishin, G.I., Sorov, Yu L., Khodataev, K.V., and L.P. Yavor, "Shock-Wave Propagation in a Decaying Plasma, Soviet Tech. Physics Letters, Vol. 8, No. 5, p. 240-241. 1982.
- [3] Lowery, H., Stepanek, C., Blanks, J., Crossway, L., M. Smith, and B. Wood, "Shock structure of a Spherical Projectile in Weakly Ionized Air," 2nd weakly ionized gases workshop, Page 273-300. Norfolk, VA. 1998.
- [4] V.P. Gordeev, A.V. Krasil'nikov, V.I. Lagutin and V.N. Otmennikov, "Experimental study of the possibility of reducing supersonic drag by employing plasma technology," *Fluid Dynamics*, Vol. 31, No. 2, 1996.

-
- [5] Robert Rubinstein, "Shock Waves in Weakly Ionized Gases," ICASE, Vol. 7, No. 1, March 1998.
 - [6] Cercignani, Carlo, *Mathematical Method in Kinetic Theory*, Plenum Press, New York, 1969.
 - [7] Chapman, S and Cowling, T. G., *The Mathematical Theory of Non-Uniform Gases*, Cambridge University Press, London, 1970
 - [8] Wang Chang, C. S. and Uhlenbeck, G. E., "On the Transport Phenomena in Rarefied Gases", *Studied in Statistical Mechanics*, 5, 1948.
 - [9] Muntz, E. P., " Rarefied Gas Dynamics", *Annual Review of Fluid Mechanics*, 21, 387-417, 1989.
 - [10] Pham-van-diep, G. C., Erwin, D. A. and Muntz, E. P., "Testing Continuum Descriptions of Low-Mach-Number Shock Structures", *J. Fluid Mechanics*, 232, 403-413, 1991.
 - [11] Fisco, K. A. and Chapman, D. R., "Hypersonic Shock Structure with Burnett Terms in the Viscous Stress and Heat Flux", AIAA paper 88-2733, 1988.
 - [12] Liaw, G. S., Guo, K. L. and Chou, L. C., "Burnett Solutions Along the Stagnation Line of a Cooled Cylinder in Low-Density Hypersonic Flows", AIAA paper 93-2726, 1993.
 - [13] L. Chou, Z.T. Deng and G.S. Liaw, "Comparison of Shockwave Structure by Solving Burnett and Boltzmann Equations," AIAA 94-2056, 1994.
 - [14] M. Zucrow, J. Hoffman, *Gas Dynamics*, Vol. 1, John Wiley and Sons, 1976.
 - [15] William M. Hilbun, "Shock Waves in Nonequilibrium Gases and Plasmas," Ph.D. dissertation, Air Force Institute of Technology, October 1997.
 - [16] Richard Seak and J. Mankowski, "Plasma Shock Wave Modification," Internal report, Accurate Automation Corporation Corp., September 17, 2001.
 - [17] Robert Rubinstein, Aaron Auslender, "Non-Ideal Gas Effects on Shock Waves in Weakly Ionized Gases," NASA CR-2000-210081, January 2000.
 - [18] C. Hirsch, *Numerical Computation of Internal and External Flows*, Volume 1, Fundamentals of Numerical Discretization, John Wiley & Sons, 1988.

Z.T. DENG is a professor in Mechanical Engineering at the Alabama A&M University. Dr. Deng instructs graduate and undergraduate mechanical engineering courses, directs graduate research, and performs research involving advanced space propulsion and computational fluid dynamics. Dr. Deng may be reached at zhengtao.deng@aamu.edu.

Biographies

XIAOQING QIAN is an associate professor in Mechanical Engineering at the Alabama A&M University. Dr. Qian instructs graduate and undergraduate mechanical engineering courses, directs graduate research, and performs research involving aerothermodynamics, plasma-dynamics and high performance computing. Dr. Qian may be reached at xiaoqing.qian@aamu.edu.

INTERACTIVE DATA VISUALIZATION AND ANALYSIS FOR MOBILE-PHONE PERFORMANCE EVALUATION

Yongsuk Lee, Florida Atlantic University (FAU), Xingquan Zhu, FAU, Abhijit Pandya, FAU, and Sam Hsu, FAU

Abstract

In this paper, the authors introduce an interactive visualization and analysis system for Drive Test Data (DTD) evaluation designed to provide first-hand mobile-phone performance assessment for different parties—including phone manufacturers and network providers—to review phone and network performance such as service coverage and voice quality. The authors propose an integrated data-visualization system, iVESTA (interactive Visualization and Evaluation System for drive Test dAta) for mobile phone drive-test data. The objective was to project high-dimensional DTD data onto well-organized web pages, such that users can visually study phone performance with respect to different factors. iVESTA employs a web-based architecture, which enables users to upload DTD and immediately visualize the test results and observe phone and network performance with respect to factors such as dropped call rate, signal quality, vehicle speed, handover and network delays. iVESTA provides a practical test environment for phone manufacturers and network service providers to perform comprehensive studies on their products from the real-world DTD.

Introduction

Drive Test Data (DTD) evaluation [1]-[5] refers to the process of evaluating mobile phone or network performance by using the data collected from the moving vehicles driving through a prearranged area with a radius of roughly 10 miles. Because DTD evaluation can provide first-hand, real-world assessments, it plays an important role for both mobile phone manufacturers and network service providers to verify the performance of their products. For example, phone manufacturers can use DTD evaluation to compare a newly-built phone with a baseline phone such that the overall performance of the new phone can be evaluated. On the other hand, for network service providers, DTD evaluation can also be used to validate the signal coverage and frequency planning. In order to provide a reliable DTD evaluation that reveals the actual phone/network performance, the test has to be performed multiple times under complex conditions in order to increase the accuracy of the test results [1], [6].

Depending on the DTD evaluation objectives and the parties actually carrying out the test, DTD data collected from the field vary significantly. For example, DTD data collected from the phone manufacturers usually have phone performance details but lack network-side information such as the status of the Mobile Switch Center or Base Stations. Due to

privacy and security issues, and the fact that mobile-phone manufacturers and network service providers may have established DTD standards, they usually do not share data with each other, which unavoidably produces a low-data integrity challenge for DTD evaluation. Low integrity means that although DTD data collected from the field can tell what happened, the data may not provide sufficient information to answer a question like why it happened, due to the limitations of the data-collection devices and the availability of network/phone status information. In summary, then, outlined here are several challenges for DTD evaluation and arguments for a DTD evaluation system that can effectively resolve all these challenges and provide a clear picture for users to answer these two questions: what happened and why did it happen?

- Establishing proper methods for evaluating a low-quality and low-integrity DTD evaluation. A DTD evaluation system must be able to decouple numerous factors such as terrain types, interference from surrounding vehicles or buildings, network variation, and phone performance.
- Deriving reliable comparisons based on a small number of significant events in DTD. Although a DTD database usually has a large volume of test data, the number of significant events, such as dropped calls, is actually very small, which makes it difficult to perform reliable statistical analyses.
- Visualizing a large number of events on the screen. Mapping DTD to a visual map can provide clear interpretation for the data, but DTDs usually have a large volume of data and events. For web-based user interfaces on which iVESTA is currently based, drawing and handling a large volume of data and providing immediate results is a challenge.
- Providing comprehensive summarized reports for DTD evaluation. DTDs usually involve a large amount of test data in which the majority of the recorded events are trivial. Consequently, it is required to provide users with comprehensive, summarized reports along with visualized presentations, such that users (possibly project managers) can gain a high-level understanding of the phone performance.

In order to resolve the above challenges, the authors propose a web-based visualization system, iVESTA, which is able to evaluate a large volume of mobile-phone DTD data

and immediately provide summaries for the users. iVESTA can carry out baseline comparisons and build standardized evaluation methods from DTD, so it can help phone manufacturers and network service providers to analyze the collected data. For example, the differences between the test phone and the baseline phone can be easily evaluated when the data are displayed on a map with signal strength, signal quality, distance between a mobile phone and a base station, different operations on a mobile phone, or network traffic at different times.

The web-based visualization architecture ensures that iVESTA can rely on a centralized DTD database and provide a variety of data analyses, summarization, and visualization functionalities. More specifically, the inherent merit of iVESTA is two-fold: (1) It provides full-scale evaluation of DTD data and is able to create baseline comparisons in call performance and RF performance; and (2) It is a dynamic visualization system, where users can easily review all existing reports, generate a new report from the log files of a specific product, and provide interactive charts and tools.

which the DTD from this study were collected, uses QPSK, 16QAM, and 64QAM digital modulation, AMBE+2 (Advanced Multi-Band Excitation), and VSELP (Vector Sum Excited Linear Predictor) speech-coding techniques coupled with TDMA (Time Division Multiple Access) channel-access methodology to enhance channel capacity and system services. It is also important to evaluate QAM modulation performance versus RF measurements (RSSI: Radio Signal Strength Indicator and SQE: Signal Quality Estimation).

RSSI in dBm is measured according to Eq. (1), where P_1 is the received signal strength and P_0 is the unit signal strength (1mW).

$$P = 10 \log(P_1 / P_0) \quad (1)$$

RSSI consists of three major components defined by Equation (2).

$$\text{RSSI} = C \text{ (Desired signal strength)} + I \text{ (Interference strength)} + N \text{ (Noise strength)}. \quad (2)$$

Table 1. Summarized influence factors of DTD: this table shows the complexity of DTD evaluation, where the mobile phone or network performance crucially depends on multi-dimensional features

Influence factors of DTD					
Network	Environment	Hardware	Software	Operator	Test
<ul style="list-style-type: none"> - Congestion - Coverage - Handoff Delay - Interference - Color code configuration 	<ul style="list-style-type: none"> - Terrain type - Natural Barriers - Artificial Barriers - Near Interference - Multi path - Weather 	<ul style="list-style-type: none"> - Stability - Hardware Fault - Heat Dissipation - Antenna Setting - Antenna Pattern - RX/TX Performance 	<ul style="list-style-type: none"> - Stability - System Fault - Mobility Management - Vocoder 	<ul style="list-style-type: none"> - Contacting - Blocking - Voice Quality 	<ul style="list-style-type: none"> - Car Type - Drive Route - Drive Direct - Antenna Extension - Speed

Consequently, iVESTA provides extensive functionalities for phone manufacturers and network providers to understand the correlation of the cause and effect on mobile performance during the drive test [10]. General influence factors of the DTD data are summarized in Table 1. The rest of the paper outlines the iDEN background to help interested readers understand the drive-test process; introduces iVESTA system architecture; provides a detailed introduction to system functionalities, with a focus on summarization tools; and, highlights the dynamic visualization components of iVESTA.

Background

A. iDEN Characteristics

iDEN [10] is a digital radio system providing integrated voice and data services to end users. The iDEN system, in

The calculation of SQE (Signal Quality Estimation) is thus given by Equation (3).

$$SQE = C / (I + N) \quad (3)$$

In iDEN systems, radio channel bandwidth assigned to 800MHz and 900MHz is 25kHz, and a single in-bound/outbound frequency pair is simultaneously shared among six users (i.e., 6 mobile radios) for dispatch voice quality and shared among 3 mobile radios for voice quality. Each time slot length is 15 milliseconds, so the communication structures of the iDEN system affect call performance during the test and require a DTD evaluation system to separate data into different call modes.

Figure 1 provides a high-level overview of the iDEN system, where DAP (Dispatch Application Processor) is responsible for the overall coordination and control of dispatch communications. An iDEN network can connect to the

PSTN through a Mobile Switching Center (MSC). The Metro Packet Switch (MPS) provides one-to-many switching between the Enhanced Base Transceiver System (EBTS) and the Dispatch Applications Processor (DAP) for dispatch voice and control, so it offers users the ability to make the Selective Dynamic Group Call (SDGC). Message Mail Service (MMS) encompasses all of the software and hardware required to store and deliver alphanumeric text messages, and the InterWorking Function (IWF) is used to manage intersystem roaming. The Base Site Controller (BSC) is the controlling element between the EBTS cell sites and the MSC that processes each type of transmission. The EBTS is the cell site that links the mobile and portable subscribers to the fixed network equipment. It is the controlling element for phone and data services. The EBTS makes it possible for subscribers to access any of the four services available in an iDEN network: telephone, dispatch, text messaging, and data.

With these dual systems, the iDEN system can be explained as a complex system, which can easily generate network delays if there is no rapid response from either the MSC or DAP system.

B. Radio Propagation Characteristics

Mobile communication is based on the propagation of the electromagnetic wave signals through the air, which follows the general laws of physics: the further the distance between the receiver and the sender, the less is the signal strength.

B.1 Signal Strength vs. Distance between the Mobile Station and the Base Station

In an open area, the power, P_o , received at a mobile-station (MS) antenna sent from a base-station (BS) antenna is given by Equation (4) [9], where P_o is the received power at the receiver, P_t is the transmitted power from the sender, λ is the wavelength, d is the distance between the sender and the receiver, g_b is the power gain of the BS antenna, and g_m is the power gain of the MS antenna.

$$P_o = P_t (\lambda / 4\pi d^2) g_b g_m \quad (4)$$

B.2 Elevated Antennas

In order to increase the mobile-station coverage area, most BS antennas are placed on the top of a cell tower—about 200 feet on average. Denoting h_b and h_m the elevated height of the base station and mobile station, respectively, the signal strength reached at the receiver is given by Equation (5) [9]:

$$P_r = P_o g_b g_m (h_b h_m / d)^2, d \gg h_b \text{ and } h_m \quad (5)$$

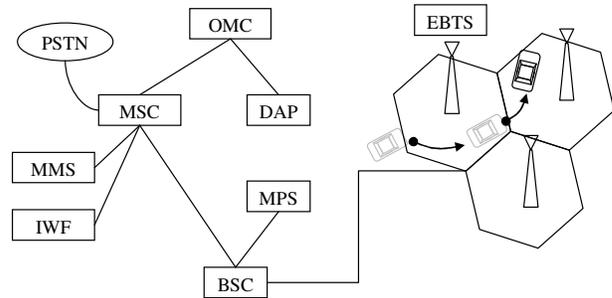


Figure 1. iDEN network system diagram

B.3 Other Factors

There are many other physical factors which have a detrimental impact on mobile communications. For instance, fading frequency is dependent on the receiver's moving speed, the reflection of the smooth objects can cause signal phase shift, diffraction can describe the modification of propagating waves when obstructed, and shadowing appears when obstacles occur in the path between the receiver and transmitter. Although DTD evaluation is assumed to consider all of these factors, in practice, the impacts of these facts are reduced through a pair of simultaneous tests: a test phone and a baseline phone are simultaneously tested in the field. Consequently, the DTD evaluation can provide evidence to indicate whether the abnormal RF signals for one particular phone are due to environmental issues, network issues, or phone issues.

C. Terrain Profile

For comprehensive evaluation purposes, the drive-test ground should cover different types of terrains. The test ground of iVESTA is shown in Figure 2, which covers a region 10 miles in diameter and includes four terrain types according to Okumura's morphology categorization [3].

- Open area: an area without any major obstacles – farmland.
- Suburban area: an area with houses, trees and low-density housing - town or small city.
- Urban area: an area with at least a two-story building - a big city.
- Mixed area: an area with one suburban side and one urban side.

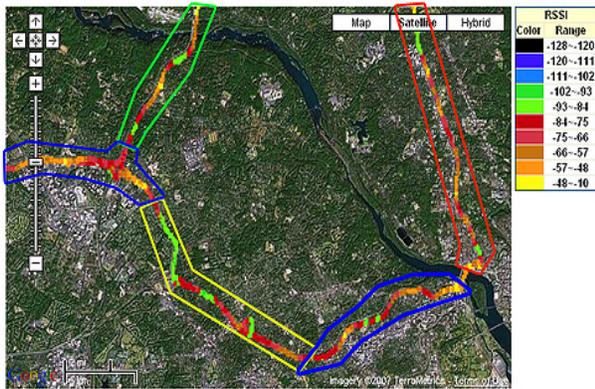


Figure 2. The four types of terrains of our drive test ground in Washington D.C (about 10 miles in diameter). Mapping tool using Google map plots RSSI data on a route in different color keys of RSSI whose table shows at the right side of the figure. The drive test area is defined into 4 different terrains: Open area in green, Suburban in red, Urban in yellow, and Mixed area in blue

D. Phone Mode Characteristics

Like many other mobile phones, iDEN phones are designed to save power, while not in use. This design generates different operation patterns for phones to listen to the BS and update its information.

D.1 Idle Mode

In the idle mode, a mobile phone will perform cell selection and reselection procedures. After a mobile phone is powered up, it performs an acquisition procedure to register to the network through a nearby base station and then listens to the paging messages broadcasted from the base station. A mobile phone in the idle mode will update its locations to the BS if necessary. Because of this, the RF measurements from scanning the neighbor cells are less accurate than the measurements of the in-call mode.

D.2 Interconnect Call vs. Dispatch Call Modes

An iDEN call mode can be separated into two types: interconnect call and dispatch. For an interconnect call, two channels—one for inbound and the other for outbound—are assigned for the call. When a call is initiated in a moving vehicle, the phone can assist the network in determining when a handover to another cell is expected. The mobile continuously monitors outbound signals from neighboring cells and measures the received power and signal quality of the signals. When the mobile determines that the neighbor measurement is higher than the measurement from the network in its currently assigned cell, it transmits a handover request to the network.

Unlike interconnect calls, which use two channels, a dispatch call uses only a single channel. The dispatch call requires the message senders and receivers to take turns sending their messages, i.e., the sender presses a button while talking and then releases the button after finishing the conversation after which the listener will press a button to make a response. In this way, the system knows the exact direction of the signal transmission.

iVESTA System Architecture

As shown in Figure 3, iVESTA mainly consists of the following ten components: drive-test log files; automatic compressor/uploader; automatic data processor; manual logs for voice defect; an integrated database; a query-driven filter/modeling; text-based summary and interactive chart; analytical mapping tools; data charts; and, replay tool.

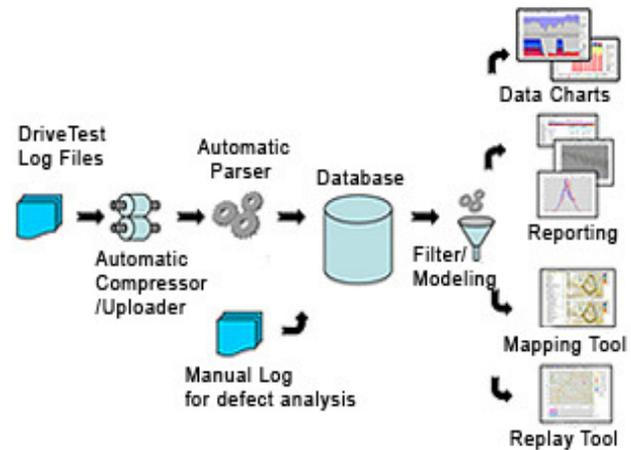


Figure 3. The main components of the iVESTA system

The structure of a log file is unique because a DTD evaluation is generated through a Radio Application Layer Protocol (RALP) device attached to the mobile phone and a computer inside a drive-test vehicle, as shown in Figure 4. In addition, iVESTA is also designed to accommodate voice quality reports, which record the voice quality of extended conversations.

To generate drive-test data, it is necessary to have a team of three people in order to operate the mobile phone, vehicle, and the base counterpart of the mobile phone. For a baseline comparison test, it is required to have a pair of two teams. Each log file mostly contains 8 hours of drive-test data, about 40 Mbytes. The cost of a drive test is expensive and the quality is poor, when compared to the in-lab simulation results. For a new product, it is required to test for several days. Therefore, web-based visualization systems play an important role in generating rapid, relevant comparison

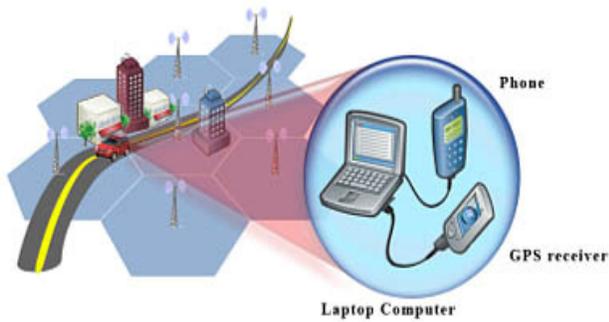


Fig. 4. The mobile drive test environment

summaries to evaluate and analyze the root cause of the defects. For the drive-test data iVESTA is dealing with, since the objective is to evaluate the new mobile phone, the baseline phone is sometimes tested with an extra antenna in order to receive a strong signal taking into consideration the vehicle penetration loss [7].

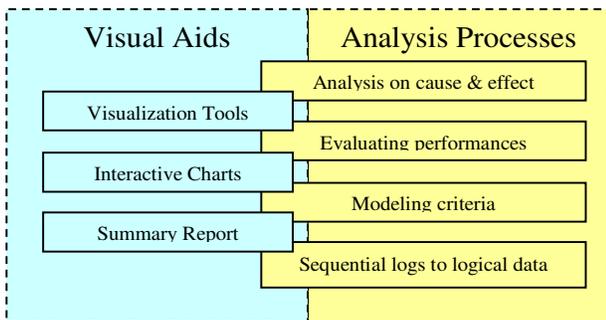


Figure 5. The Architecture of iVESTA

While the iVESTA system is intended to offer full support for the DTD evaluation, without cell site information and a channel plan from the network service providers, it is difficult to make unbiased comparisons and conclude the phone's performance. To overcome this low-data integrity hurdle, iVESTA employs a set of visual aids to bridge the gap between different levels of data understanding. As shown in Figure 5, the right side of the picture shows four layers of data analysis: (1) converting sequential logs into logical data; (2) modeling criteria; (3) performance evaluating; and (4) cause-and-effect analysis. In order to bridge the gap between two consecutive data analysis layers, a set of visual aids are employed to help users understand the data and characterize the phone's performance, such that the low-data integrity challenge can be resolved in practice.

A. DTD Format

RALP logs are sequential messages between a Mobile Station (MS) and the Enhanced Base Transceiver System (EBTS), beginning with PC time stamp and a tick count to

establish ordering of additional messages within a millisecond, followed by an information log specifying the phone and conversation status. Figure 6 demonstrates the basic elements of a DTD record. Because all records and data fields are well-defined, an automatic data processor can be employed to read each single record and pull corresponding data into the database.

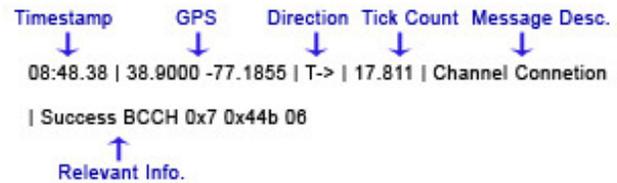


Figure 6. An example of a DTD log

B. Log Files Uploading

After DTD logs are collected by a drive test team, the files are uploaded through a web-based uploading tool as shown in Figure 7, which consists of three major parts: (1) a user information section, (2) an uploading section for DTD logs of one particular product, and (3) the uploading section for DTD logs of the baseline product. Log files are guided with proper naming conventions procedure for avoiding duplicates. In addition, all log files are automatically compressed on the local machines and sent to the server through the HTTP protocol.

The uploading process triggers iVESTA's automatic data processor to read log files and populates data into the main database (powered by MySQL), which consist of seven tables: NMS, Calls, Mobility, Events, LAPidata, Site, and Drivetest. The uploaded information, such as transaction ID and the uploader's information, is stored into the upload table. The drive-test table is a look-up table to synchronize the relationships between database tables and the log files.

The Normal Mode Summary (NMS) table contains the main DTD, such as geographical information and speed for vehicle and cell tower, PC timestamp, server color code, RSSI, SQE and measurement, the first neighbor's RSSI, SQE and measurement, the second neighbor's RSSI, SQE

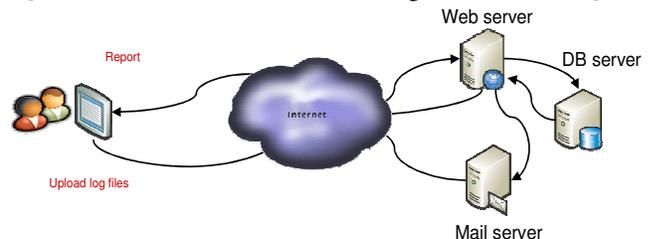


Figure 7. The upload process for log files

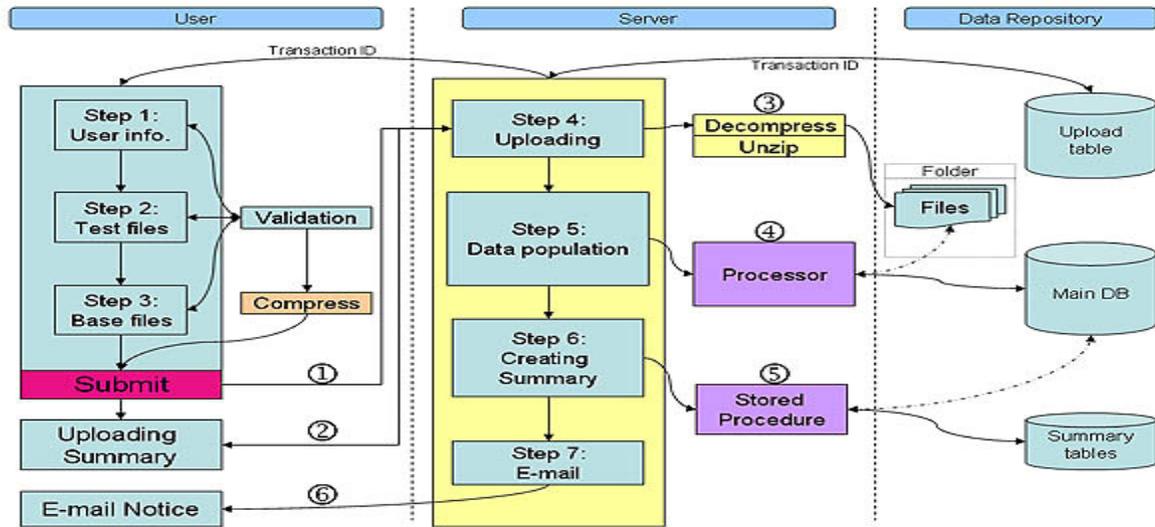


Figure 8. The data process diagram of iVESTA log file uploading

and measurement, and so on. The calls table is populated with call information such as call type, total call length and so on. The mobility table has mobility information such as the cause of handover decision, oldSitePCCH, newSitePCCH, and so on. All of the sequential event messages between MS and BS are found in the events table. Packet data information is stored in the LAPidata table. The site table places cell information such as GPS information of cell, server color code, and so on.

Following the data conversion, a summary process will create a set of charts, which may be used to help users generate a series of reports. Upon finishing the file uploading, the mail server will notify the drive-test team and their clients through an email message. A detailed diagram of the data process for uploading log files is shown in Figure 8. The uploading process starts from ① at which point a user submits the log files, which consequently triggers processing, and ② returning uploading confirmation and summary to the user. In step ③, the system decompresses the uploaded files, followed by step ④, which automatically process the log data and pulls useful information into the main database. In step ⑤, the system creates a set of summary tables from the input data. Finally, then, in step ⑥, the system sends an email confirmation to the submitter and the system administrator to acknowledge the success of uploading the test data.

C. Summary Reports

The objective of the summary reports for the DTD evaluation is to provide users the statistical overview of a drive-test

report: the informative summaries of events, the total calls in different modes, the average signal conditions on a route, and so on. The summary reports provided by iVESTA contain 8 main performance evaluations, which are 22 text-based summary tables, as shown in Figure 9. Users can access all of these tables by navigating the system submenus (the left panel in Figure 5.)

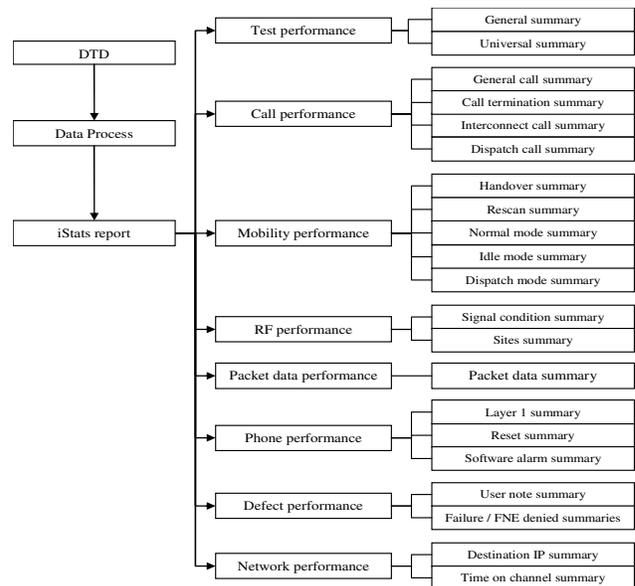


Figure 9. iVESTA performance summary diagram

In short, the summary tables cover two important components: (1) the baseline comparison form and (2) the statistical calculation based on the selected comparison attributes

such as the total number of dropped calls between the test and the baseline phones. From the phone manufacturers' perspective, the most interesting parameters are listed as follows:

- Connected call percentage: number of calls connected to TCH (traffic channel) over number attempted.
- Good call percentage: number of calls ended properly over number of calls connected to TCH.
- Number of rapid handovers: two handover attempts within 10 seconds.
- Number of delayed handovers: MR sent but no handover command received within 5 seconds of the first MR sent.
- Number of Ping-Pong handovers: two handover attempts, where the radio goes from cell site A to cell site B and back to cell site A.

To provide statistical comparison analyses, iVESTA employs a two-proportion Z test with a default null hypothesis ($p\text{-value} > 0.05$) in all summary tables. The diagram of a two-proportion Z test is shown in Figure 10, and Figure 11 pictorially shows a summary table with statistical test results.

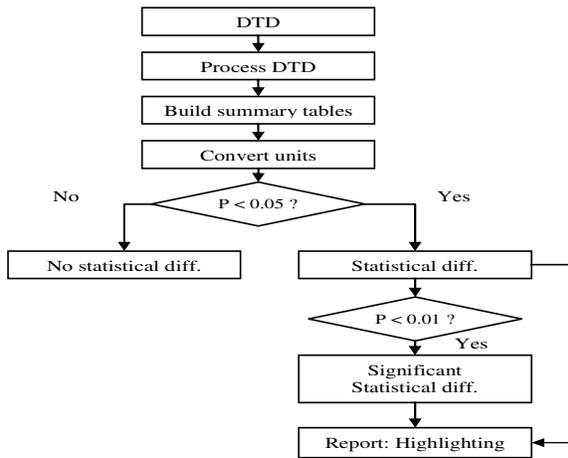


Figure 10. The diagram of a two-proportion Z test of the summary tables

D. Interactive Charts

Evaluating RF measurements of a DTD evaluation is the most important assessment in iVESTA because about 40% of abnormal events, such as dropped calls, occur in poor-signal conditions (less than -100dBm). iVESTA provides regional comparisons of RF measurements in different call modes since signal condition strongly relies on the distance and terrain type between MS and BS. In addition, interactive

charts also provide spontaneous responses corresponding to a mouse position on a chart, which is shown in Figure 12.

iVESTA provides eight groups of 40 user-interactive charts and is able to let a user visually evaluate the phone performance with respect to different parameter settings. For example, the SQE (Signal Quality Estimation) charts are drawn into CDF (cumulative distribution function) and PDF (probability density function) forms and separated by five

iStats Summary					Print or Save
Search by					
Test SW ID: D2F.01.31	Tested Date(s): 05/04/2007, 05/07/2007, 05/11/2007				
Baseline SW ID: D3D.01.13	Tested Date(s): 05/07/2007, 05/08/2007, 05/09/2007				
Values normalized to: 100 min. 0 sec. Confidence Interval: <input type="checkbox"/> % <input type="button" value="Apply"/>					
If P-value < 0.05, Raw(count) cell in this color.					
If P-value < 0.01, Raw(count) cell in this color.					
	D2F.01.31	D3D.01.13			
	Raw (Count)	Raw (Count)			BASE
Failed Handovers	0 Ea.	0.00	0 Ea.	0.00	Phone
Incomplete Handovers	0 Ea.	0.00	0 Ea.	0.00	Phone
Rapid Handovers (< 10 sec. between HOs)	1 Ea.	0.23	8 Ea.	1.73	Phone
Delay Handovers (> 5 sec. between HOs)	0 Ea.	0.89	82 Ea.	17.76	Phone
Ping Pong Handovers	0 Ea.	2.95	32 Ea.	6.93	Phone
Assignment Command	3 Ea.	3.40	18 Ea.	3.90	Phone
Assignment Access	3 Ea.	3.40	18 Ea.	3.90	Phone
Successful Assignment	3 Ea.	3.40	18 Ea.	3.90	Phone
Failed Assignment	0 Ea.	0.00	0 Ea.	0.00	Phone
Measurement Inquiry	0 Ea.	0.44	68 Ea.	14.73	Phone
Coverage Test	0 Ea.	0.00	0 Ea.	0.00	Phone
Trace	0 Ea.	0.00	0 Ea.	0.00	Phone
Congestion Relief	28 Ea.	6.35	13 Ea.	2.82	Phone
Poor Uplink	18 Ea.	4.08	55 Ea.	11.91	Phone
---Poor Uplink RSSI > 90, SQE > 15	9 Ea.	2.04	5 Ea.	1.08	Phone
Intelligent System Optimization	0 Ea.	0.00	0 Ea.	0.00	Phone
Timed Out MIs	0 Ea.	0.00	0 Ea.	0.00	Phone
Measurement Reports	340 Ea.	77.15	521 Ea.	112.83	Phone
MR - MI	294 Ea.	66.71	453 Ea.	98.11	Phone
Ratio of Assignment to MI	0.33	1.00	0.26	1.00	----

Draggable P-value table

P-value (Two proportion z test)

Hot/Cold Hypothesis: P-value >= 0.05

# of Intercon.-call Records	Rapid Handovers (< 10 sec. between HOs)
Test: 23378	Test: 1
Baseline: 22761	Baseline: 8
P-value: 0.0188	
Decision: Statistical Difference	

Figure 11. An example of a text summary table with a statistical information panel

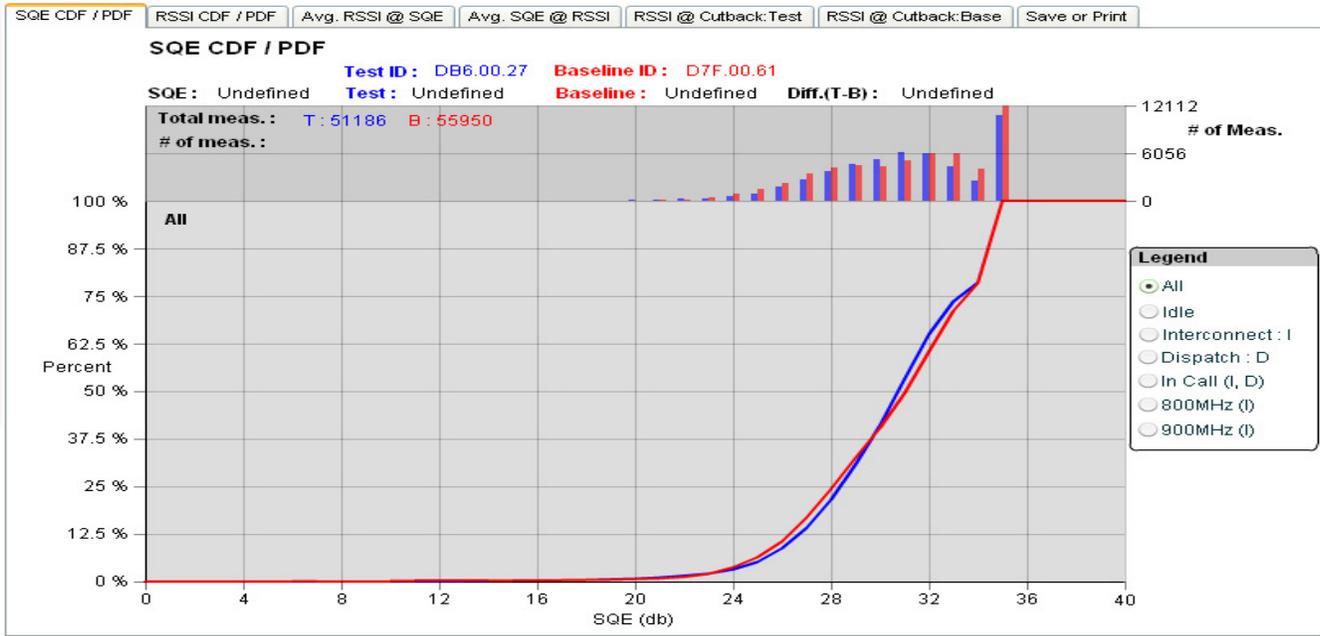


Figure 12. Interactive chart interface: all interactive charts can be navigated by selecting the menu on the left side of the figure. The lower curve graph in blue represents SQE CDF of a product under test, and the upper graph in red represents SQE CDF of a baseline product. The x-axis indicates SQE in the range from -10dB to 40 dB, and the y-axis represents the cumulative percentage in the range from 0 to 100.

different phone modes: all, in-call, dispatch, idle, and interconnect call. The interactive charts generally describe the differences between signal conditions and then evaluate the performance of the test phone with respect to the baseline phone with a particular signal condition. More specifically, if the SQE graph of a test phone is lower than a baseline phone in CDF, it means that the test phone has received a higher signal quality than the baseline phone.

E. Voice and Packet Data Evaluation Report

In addition to the evaluation at the event level, iVESTA can also provide voice and packet data evaluation reports. Consequently, the users can understand the phone performance with respect to the voice and the data transmission qualities.

During a drive test, the voice communications over mobile phones are recorded by the drive test team. To find a correlation between voice defects (unrecognized voice conversation), phone events and RF measurements, a manual process is required to compare the recorded conversation during the drive test and then store the labeled information in the defect database, as shown in Figure 13. The voice defect report is evaluated through the comparison between

the test phone and the baseline phone. An example of the voice defect chart is shown in Figure 14.

Since a call communication between two users is based on the audio, voice quality is a very important factor that should be properly evaluated. Currently, iVESTA can evaluate the following types of defects: busy, choppy audio, clicks and clunks, customer unavailable, dropped call, echo, extended-muting, fading, FNE denied, freeze, garble, helicopter, iDen freeze, mute, no audio, no connect, NT, one way audio, other, and out-of-service and special codes for defects. A sample of the voice defect chat is shown in Figure 14, where the x-axis indicates the type of defects, and the y-axis represents the percentage of a defect over the total number of calls. The bars in the column indicate the defects of a product under test and baseline product. The highlighted defect in Figure 14 shows that 1.25% of calls on the baseline product have “no audio” defect, while the product under test has 0.59% only, which is clear evidence that the test phone outperforms the baseline phone with respect to the voice defects.

For a packet-switch network, the voice channels and dispatch functions are used to transmit packet data other than voice, such as emails. Due to the sharing of voice channels,

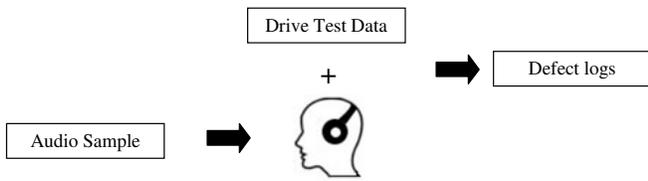


Figure 13. The diagram of voice defect labeling process: a manual process is employed to evaluate the phone conversations recorded from the drive test, and the labeled voice defects are stored in a defect database.

the actual data rate and flow will depend on the network traffic. Like iDEN, WiDEN allows compatible subscriber units to communicate across four 25 kHz channels combined, for up to 100 kbit/s of bandwidth. According to the iDEN characteristics [10], four carrier allocations for each 25 kHz bandwidth should be evaluated. For this purpose, the inbound and outbound throughputs with respect to the RF measurements are employed to indicate the packet data transmission quality. The LAPi data time chart is designed to visualize the carrier allocation performance with respect to the inbound and outbound throughputs and RF measurements. Figure 15 shows a packet data performance chart on a timeline that is placed in the middle of the chart. RSSI and SQE are shown at the top of the chart. At the bottom of the chart, four WiDEN carrier allocations and one iDEN carrier allocation are depicted with respect to the SQE timeline. This chart provides interactive and scrollable panels for test and baseline phone. From Figure 15, one can see that WiDEN carrier allocations became lower and move to iDEN carriers when the values of SQE became lower.

iVESTA Visualization Tools

The development of visualization tools for DTD evaluation is becoming an imperative demand for mobile-phone manufacturers and network providers because such tools help them analyze the collected data directly related to the customer environments. In practice, such data only make sense when interpreted in high dimensions. For example, the differences between test and baseline phones can be easily evaluated when the collected data are displayed on a map with signal strength, signal quality, distance between mobile phones and base stations, and network traffic. While the summarization tools introduced previously may indicate what happened during the drive test, e.g., dropped calls, the objective of the iVESTA visualization tools is to clearly answer why it happened, i.e., what caused the low RSSI and why the network did not properly hand over the phone to a neighbor cell.

A. Mapping Tool

The iVESTA mapping tool is a web application with selective features such as the test dates, RF measurements, and events on a map for different phones (test vs. baseline phones), as shown in Figure 16. Users can select different test dates for both phones and call modes. For all events recorded in the DTD database, some of them, such as dropped calls or rescans, are only understandable with RF measurements.

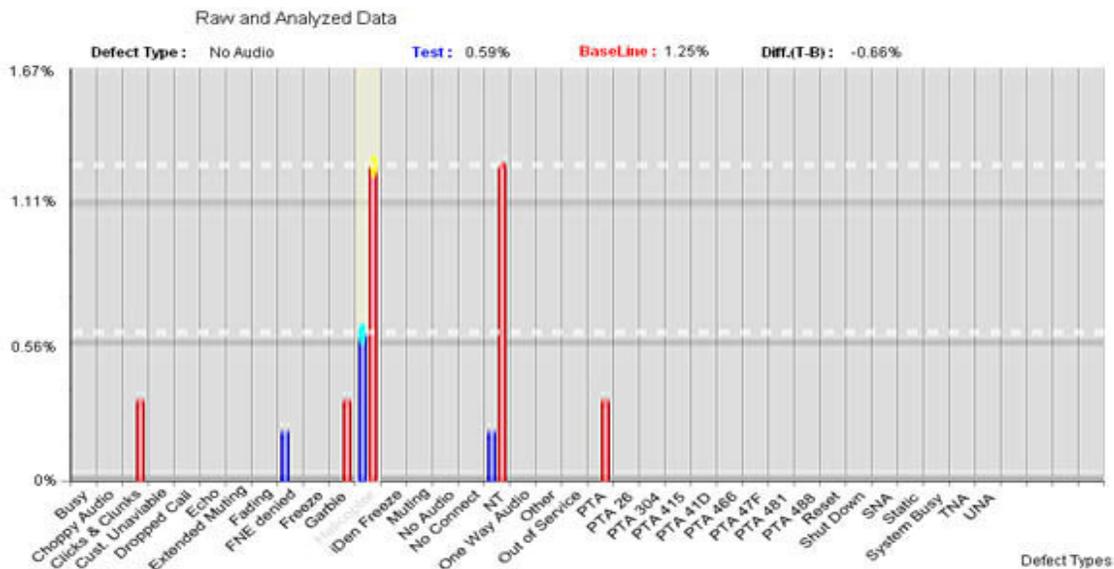


Figure 14. A sample voice defect chart

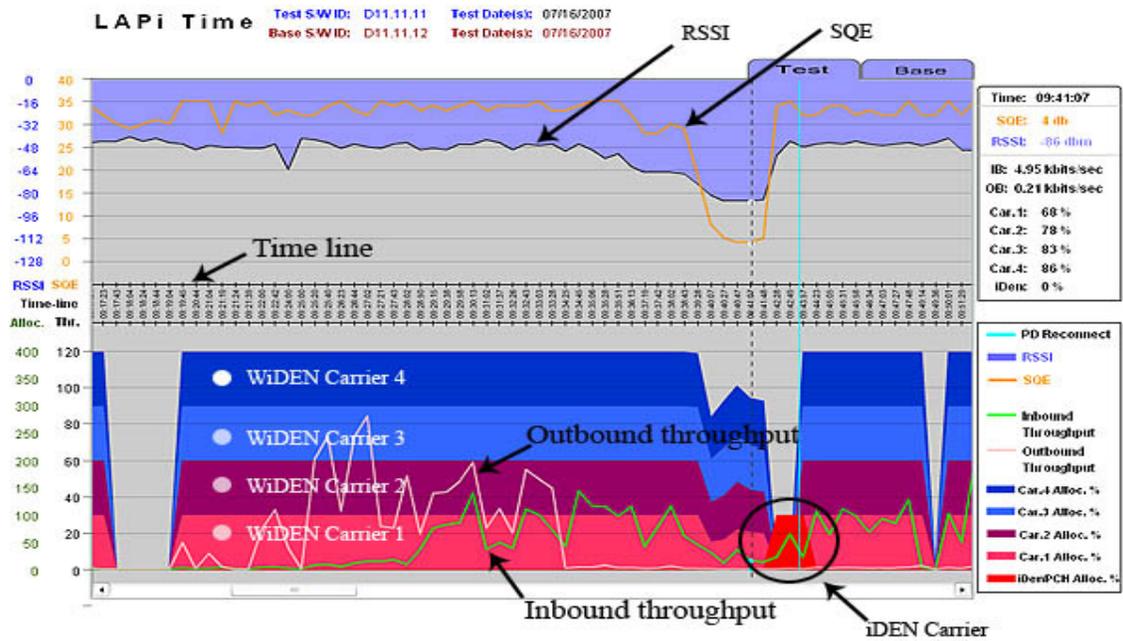


Figure 15. Sample LAPi time chart

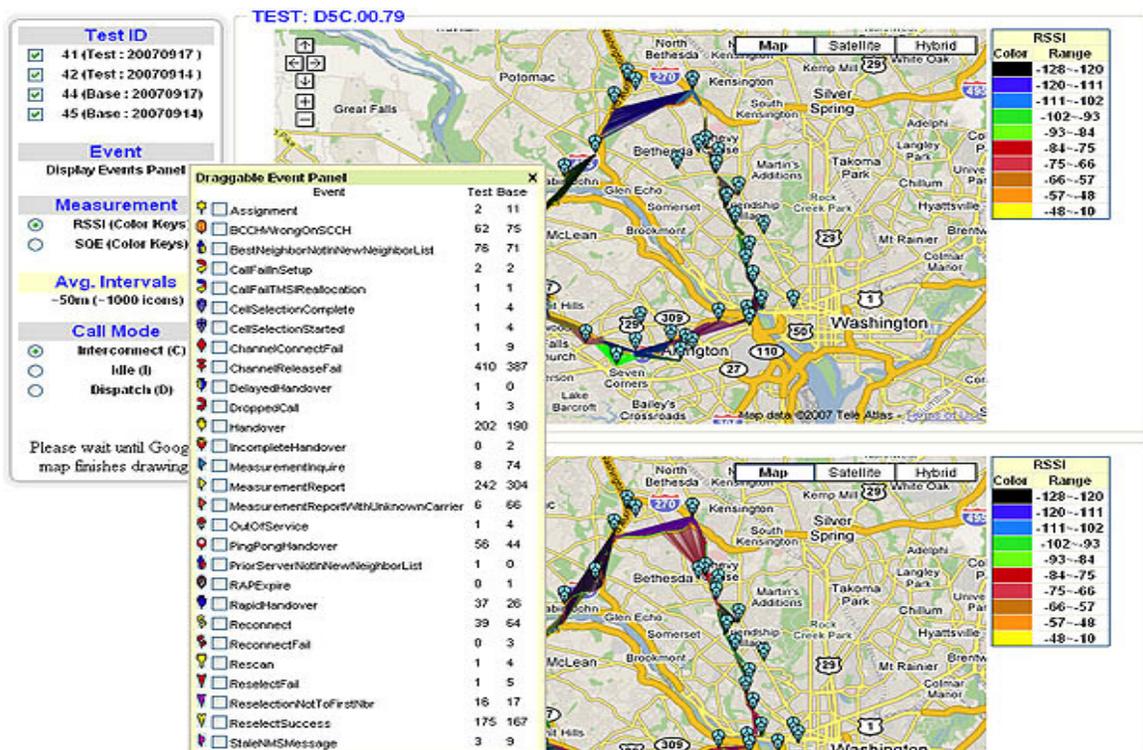


Figure 16. The mapping tool with various features: two dynamic Google maps for test and baseline phones, and moveable event panel, menu with selectable options, and icons with detailed event information.

In the iVESTA system, the RF measurements such as RSSI or SQE are visualized with color keys corresponding to the ranges of the measurements. In addition, users can selectively decide which event(s) should be displayed such that the users can review all event details. Another sample of the mapping tool is shown in Figure 17, which displays all towers and their links to the route where a mobile phone is present. In order to provide a web-based dynamic map, this tool is designed with Google map API. Instead of reviewing the coverage of the network service, this tool helps users to clearly observe abnormal mobility performances.

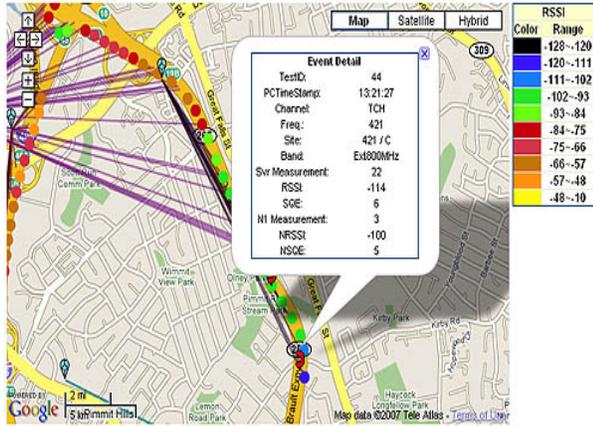


Figure 17. The detail information of a dropped call event with zoom-in state: the event detail window contains Test ID, PC time stamp, PCCH frequency and color code and RF measurements on server and foreground neighbor.

In order to reduce the time of drawing all icons on the Google map, the mapping tool is displayed in a 2-layer structure: the transparent layer is used for RF measurements and the icon layer is designed for the events. The data processing flowchart for the mapping tool is shown in Figure 18. Whenever the Google event listener catches a user event, it calls Ajax functions to retrieve data from the server and renders the map with the event icons.

B. Replay Tool

Different from the mapping tools that simply map the events to the web-pages, the replay tool helps users restore drive-test scenes with respect to factors such as the directionality and the velocity of the drive-test vehicles, through an animated web page, which dynamically shows the drive-test trail, as shown in Figure 19. In addition, the RF measurement is also dynamically visualized such that the users can understand the test environments such as signal strength, directions, handovers, and cell towers to which the phone was connecting; the bottom part of Figure 19(a) displays the

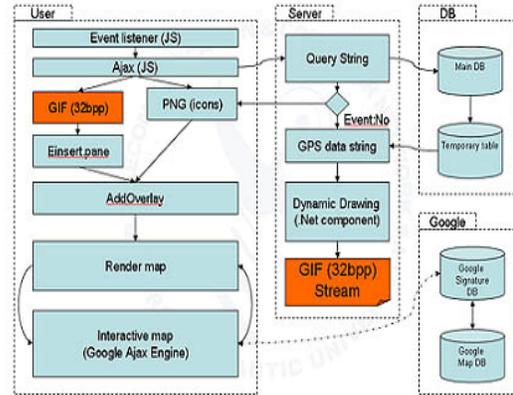
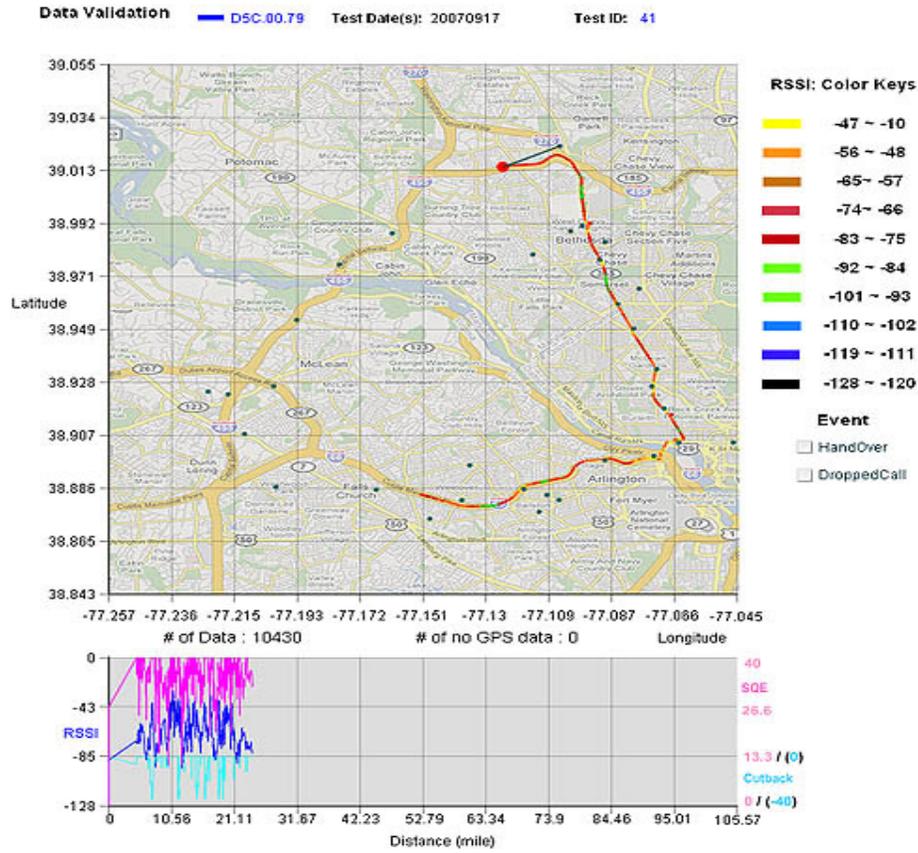


Figure 18. The data process on the mapping tool.

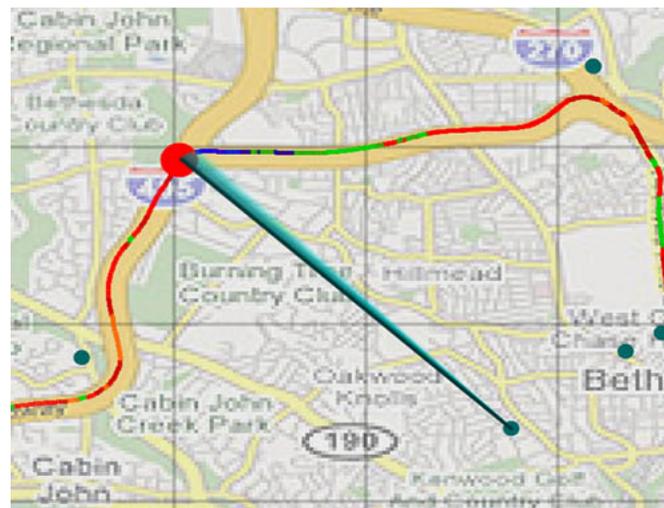
RSSI and SQE measurements along the drive-test trip. The directionality of the test route is a critical factor to the mobility evaluation because the zones of the mobility evaluation are very sensitive and dependent on this factor. At the beginning of the development of the iVESTA system, a large portion of the distinct handovers between the test and the baseline phone were observed. Later, this issue was analyzed as a cause of the vehicle directionality. This tool animates a drive-test vehicle with link lines between an MS and a BS, so it helps users to review the coverage of the network services.

Summary

Provided here was a high-level overview of the iVESTA system, which was developed for drive test-data evaluation with the objective of providing generic functionalities to help a drive-test team analyze DTD evaluations and understand the mobile-phone performance in comparison with the baseline products. iVESTA delivered a practical product for DTD evaluation, which successfully resolves the low-integrity, low-quality, high-uncertainty, and low-interpretability issues of the DTD. In short, iVESTA is a web-based reporting system which provides a full-scale evaluation of DTD through extensive comparison with the baseline phone. In addition, iVESTA provides numerous visualization tools to help users understand their DTD and essentially resolve the concerns about what happened and why it happened. This study focused mainly on the DTD evaluations from the mobile phone manufacturers' perspective. To extend this research, more efforts on scalability for a large number of subscriber clients and the extension for network service providers are highly recommended. In addition, the authors plan to focus on providing automatic scoring models for mobility and RS measurements, seamless access from chart to DTD, analysis on the cause and effect, and integrating data mining tools [8] to support intelligent data analysis in a subsequent study.



(a)



(b)

Figure 19. A screen shot of the replay tool: (a) the top part of tool illustrates the path of drive test vehicle on a map and at the same time displays the RSSI in different color keys. The bottom part of the tool plots RSSI, SQE and cutback measurements simultaneously. (b) a detailed view of the replay tool: the links between MS and BS are traced.

Acknowledgement

This research was partly supported by a grant from Motorola to Drs. Pandya and Hsu.

References

- [1] Y. Lee, X. Zhu, A. Pandya, and S. Hsu, iVESTA: an Interactive Visualization and Evaluation System for Drive Test Data, in *Proc. of the ACM Symposium on Applied Computing*, pp.1953-1957, Brazil, 2008.
- [2] R. Miletic, Using Autonomous Drive Test Data to Improve Network Quality and Engineering Efficiencies, in *Proc. of the World Wireless Congress annual Conference*, 2006.
- [3] Aircom Worldwide, 3G Live Network Optimised using Combined ACP and Drive Test Data, <http://www.aircominternational.com/Default.aspx?locID=04vnew008.htm>
- [4] M. J. Flanagan, L. M. Drabeck, L. A. Cohen, A. H. Diaz, J. Srinivasan, Wireless network analysis using per call measurement data, *Bell Labs Technical Journal*, 11(4):307-313, 2007.
- [5] Wu Jiao, Huang Yiling, Wei Zaixue, and Yang Dacheng, A Novel Model of Drive Test Data Processing in Wireless Network Optimization, in *Proc. of the 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 2006.
- [6] Jianming Zhang, Juanjuan Sun, and Dacheng Yang, Application of Drive Test for QoS Evaluation in 3G Wireless Networks, in *Proc. of ICCT*, 2003.
- [7] Kostanic, I.; Hall, C.; McCarthy, J., "Measurements of the vehicle penetration loss characteristics at 800MHz", Vehicular Technology Conference, 1998. VTC 98. 48th IEEE.
- [8] Xingquan Zhu and Ian Davidson, *Knowledge Discovery and Data Mining: Challenges and Realities*, IGI publishing, April 2007.
- [9] D. O. Reudink, Properties of mobile radio propagation above 400 MHz, in *Proc. of the 24th IEEE Vehicular Technology Conference*, Volume 24, Dec. 1973 P61 – 79.
- [10] iDEN, Motorola iDEN Technical Overview, <http://webaugur.com/matt/files/nextel/techover.pdf>

Biographies

YONGSUK LEE received his B.S. and M.S. degrees in computer science from Florida Atlantic University (FAU), Florida, and Korea in 2005 and 2007, respectively. He joined Motorola, Florida, in 2005 and is currently working as a programmer analyst in the Drive Test Team for iDEN

Technology Division. His current research areas of interest include data analysis on WiMAX technology, data visualization, and real-time web technology.

XINGQUAN ZHU is an Assistant Professor at the Department of Computer Science and Engineering, Florida Atlantic University, Boca Raton, FL. He received his Ph.D degree in Computer Science from Fudan University, Shanghai, China, in 2001. From Feb. 2001 to Oct. 2002, he was a Postdoctoral Associate in the Department of Computer Science, Purdue University, West Lafayette, IN. From Oct. 2002 to July 2006, He was a Research Assistant Professor in the Department of Computer Science, University of Vermont, Burlington, VT. His research interests include data mining, machine learning, multimedia systems, and information retrieval. Since 2000, Dr. Zhu has published extensively, including over 60 refereed papers in various journals and conference proceedings.

ABHIJIT PANDYA is a professor at the Department of Computer Science and Engineering Department, Florida Atlantic University. He received his undergraduate education at the Indian Institute of Technology, Bombay. He earned his M.S. and Ph.D. in Computer Science from the Syracuse University, New York. He has worked as a visiting Professor in various countries including Japan, Korea, India, etc. His areas of research include VLSI implementable algorithms, Applications of AI and Image analysis in Medicine, Financial Forecasting using Neural Networks. Dr. Pandya has published over 150 papers and book chapters, and a number of books in the areas of neural networks and ATM networks. This includes a text published by CRC Press and IEEE Press entitled "Pattern Recognition using Neural Networks in C++". He consults for several industries including IBM, Motorola, Coulter industries and the U.S. Patent Office.

SAM HSU is a Professor at the Department of Computer Science and Engineering, Florida Atlantic University. He received his Ph.D degree in Computer Engineering in 1993 from Florida Atlantic University His research interests include Web Technologies and Computer Networking. He has published over 50 refereed journal/conference papers in these areas. He also has close ties to local industries engaging in joint projects that have practical applications. Dr. Hsu is a member of ACM, IEEE, and Upsilon Pi Epsilon Honor Society.

INSTRUCTIONS FOR AUTHORS

MANUSCRIPT REQUIREMENTS

The INTERNATIONAL JOURNAL OF MODERN ENGINEERING is an online/print publication, designed for Engineering, Engineering Technology, and Industrial Technology professionals. All submissions to this journal, submission of manuscripts, peer-reviews of submitted documents, requested editing changes, notification of acceptance or rejection, and final publication of accepted manuscripts will be handled electronically.

All manuscripts must be submitted electronically. Manuscripts submitted to the International Journal of Modern Engineering must be prepared in Microsoft Word 98 or higher (.doc) with all pictures, jpg's, gif's and pdf's included in the body of the paper. All communications must be conducted via e-mail to the manuscript editor at philipw@bgsu.edu with a copy to the editor at editor@ijme.us

The editorial staff of the International Journal of Modern Engineering reserves the right to format and edit any submitted word document in order to meet publication standards of the journal.

1. Word Document Page Setup: Top = 1", Bottom = 1", Left = 1.25", and Right = 1.25". This is the default setting for Microsoft Word. Do Not Use Headers or Footers.
2. Text Justification: Submit all text as "LEFT JUSTIFIED" with No paragraph indentation.
3. Page Breaks: No page breaks are to be inserted in your document.
4. Font Style: Use 11-point Times New Roman throughout the paper except where indicated otherwise.
5. Image Resolution: Images should be 96 dpi, and not larger than 460 X 345 Pixels.
6. Images: All images should be included in the body of the paper (.jpg or .gif format preferred).
7. Paper Title: Center at the top with 18-point Times New Roman (Bold).
8. Author and Affiliation: Use 12-point Times New Roman. Leave one blank line between the Title and the "Author and Affiliation" section. List on consecutive lines: the Author's name and the Author's Affiliation. If there are two authors follow the above guidelines by adding one space below the first listed author and repeat the process. If there are more than two authors, add one line below the last listed author and repeat the same procedure. Do not create a table or text box and place the "Author and Affiliation" information horizontally.
9. Body of the Paper: Use 11-point Times New Roman. Leave one blank line between the "Author's Affiliation" section and the body of the paper. Use a one-column format with left justification. Please do not use spaces between paragraphs and use 0.5" indentation as a break between paragraphs.
10. Abstracts: Abstracts are required. Use 11-point Times New Roman Italic. Limit abstracts to 250 words or less.
11. Headings: Headings are not required but can be included. Use 11-point Times New Roman (ALL CAPS AND BOLD). Leave one blank line between the heading and body of the paper.
12. Page Numbering: The pages should not be numbered.
13. Bibliographical Information: Leave one blank line between the body of the paper and the bibliographical information. The referencing preference is to list and number each reference when referring to them in the text (e.g. [2]), type the corresponding reference number inside of bracket [1]. Consider each citation as a separate paragraph, using a standard paragraph break between each citation. Do not use the End-Page Reference utility in Microsoft Word. You must manually place references in the body of the text. Use font size 11 Times New Roman.
14. Tables and Figures: Center all tables with the caption placed one space above the table and centered. Center all figures with the caption placed one space below the figure and centered.
15. Page limit: Submitted article should not be more than 15 pages.

College of Engineering, Technology, and Architecture

University of Hartford



For more information please visit us at
www.hartford.edu/ceta

For more information on undergraduate programs please contact Kelly Cofiell at cetainfo@hartford.edu.

For more information on Graduate programs please contact Laurie Grandstrand at grandstran@hartford.edu.

Toll Free: 1-800-766-4024
Fax: 1-800-768-5073



DEGREES OFFERED :

ENGINEERING UNDERGRADUATE

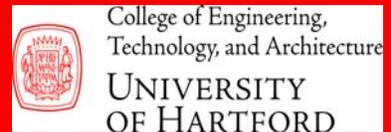
Acoustical Engineering and Music (B.S.E)
Biomedical Engineering (B.S.E)
Civil Engineering (B.S.C.E)
-Environmental Engineering Concentration
Environment Engineering (B.S.E)
Computer Engineering (B.S.Comp.E.)
Electrical Engineering (B.S.E.E.)
Mechanical Engineering (B.S.M.E.)
- Concentrations in Acoustics and
- Concentrations in Manufacturing

TECHNOLOGY UNDERGRADUATE

Architectural Engineering Technology (B.S.)
Audio Engineering Technology (B.S.)
Computer Engineering Technology (B.S.)
Electronic Engineering Technology (B.S.)
-Concentrations in Networking/
Communications and Mechatronics
Mechanical Engineering Technology (B.S.)

GRADUATE

Master of Architecture (M.Arch)
Master of Engineering (M.Eng)
• Civil Engineering
• Electrical Engineering
• Environmental Engineering
• Mechanical Engineering
- Manufacturing Engineering
- Turbomachinery
3+2 Program (Bachelor of Science and
Master of Engineering Degrees)
E²M Program (Master of Engineering and
Master of Business Administration)



IJME BEGINS ITS SECOND DECADE OF SERVICE

**IJME IS THE OFFICAL AND FLAGSHIP JOURNAL OF THE
INTERNATIONAL ASSOCIATION OF JOURNALS AND CONFERENCE (IAJC)**

www.iajc.org



The International Journal of Modern Engineering (IJME) is a highly-selective, peer-reviewed journal covering topics that appeal to a broad readership of various branches of engineering and related technologies. IJME is steered by the IAJC distinguished board of directors and is supported by an international review board consisting of prominent individuals representing many well-known universities, colleges, and corporations in the United States and abroad.

IJME Contact Information

General questions or inquiries about sponsorship of the journal should be directed to:

Mark Rajai, Ph.D.

Editor-in-Chief

Office: (818) 677-5003

Email: editor@ijme.us

Department of Manufacturing Systems Engineering & Management

California State University-Northridge

1811 Nordhoff St.

Northridge, CA 91330